Privacy-Preserving Design and Operation of Medical (Research) Registers

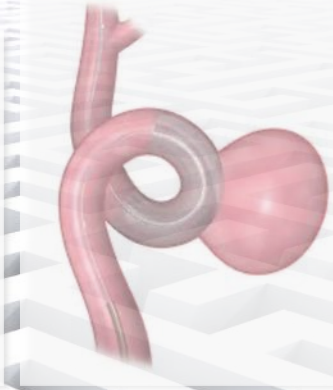
Prof. Dr. Hannes Federrath
Sicherheit in verteilten Systemen (SVS)
http://svs.informatik.uni-hamburg.de

Agenda: Privacy-Preserving Design and Operation of Medical (Research) Registers

- Introduction: Classical approach to the realization of Medical Research Registers
- IDOMENEO approach for a Peripheral Artery Disease (PAD) Research Register
  - Architecture and Workflow
  - Selected Challenges
- PANDA approach for device studies (stents, balloons) in neuroradiology (strokes and aneurysms)
  - Decentralized storage, ready for (federated) machine learning approaches
  - Prepare for upcoming Research Data Law (data usage without informed consent)
- Conclusions and future research directions

Gemeinsamer Bundesausschuss

Bundesministerium für Wirtschaft und Energie



Privacy-Preserving Design and Operation of Medical (Research) Registers

- Requirements
  - Strong Admission Control and Access Control Mechanisms
  - Encrypted Transmission of Medical Data
  - Personal Data remain in Medical Centers (Hospitals)
  - Decentralized Storage and Linking of Data Sets

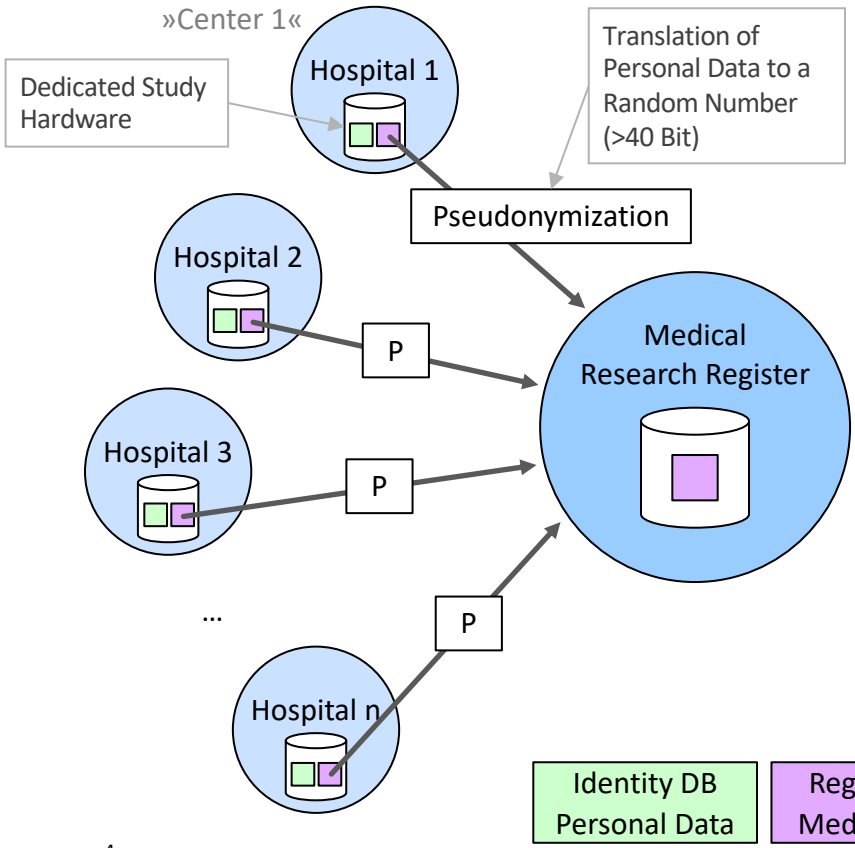
Task: Realization of a medical Register for longitudinal studies of minimum 10 years.

- Realization
  - Isolation of Technical Components within the Register
  - Communication to and from the Register via dedicated interfaces only

- Design Methods
  - Security & Privacy by Design
  - Privacy by Default

Scope: Exemplary realization of requirements coming from EU General Data Protection Regulation (GDPR)

Classical approach to the realization of Medical Research Registers



- Large number of Centers (Hospitals) allows an increased number of cases into the Research Data
- Informed Consent as a legal requirement for the inclusion of Study Participants (Patients)

- Centers need to run dedicated Study Hardware
- Some Centers prohibit connecting Study Hardware to local hospital network (transfer data via USB drives or CD-R/W)
- Operation and Maintenance (Updates, Malware protection) difficult to realize (longitudinal studies of > 10 years)

Zero-Day-Exploit Market Example

Zero-Day-Exploit Market Example

[private]

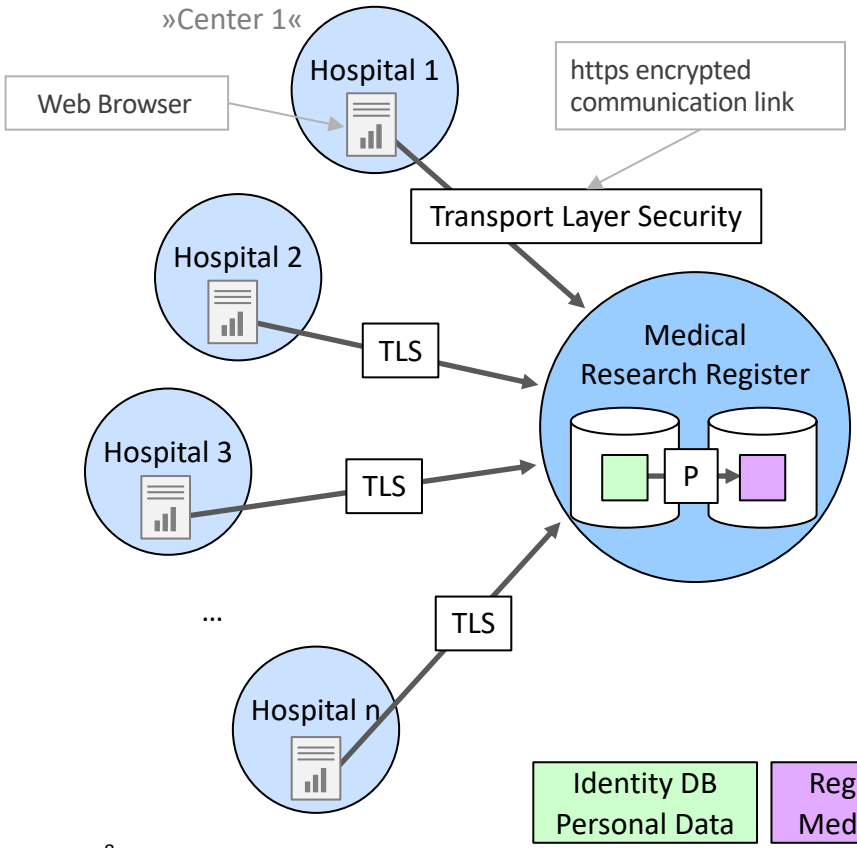
-::DATE	-::DESCRIPTION	-::TYPE	-::HITS	-::RISK				-::GOLD	-::AUTHOR
2013-04-08	Reallyeasycart 2.8.x Remote Code Execute Vulnerability	php	594		R	D	✓	10	n0tch
2013-03-07	Drupal LiteCommerce (v.lc3-1.1.3) <= Multiple Vulnerabilities	php	2483		R	D	✓	3	KedAns-Dz
2013-03-06	vShare<=2.8.1 SQL Injection + Remote Command Execution	php	2053		R	D	⚠	25	GoldenEgg
2013-02-27	uTorrent 3.x app exploit 0day	windows	4083		R	D	⚠	99	_null_
2013-02-19	vBulletin 5.0.0 Beta Release 0day Exploit	php	8392		R	D	✓	285	x00F
2013-02-17	ARASTAR Sql Injection Vulnerability	php	1857		R	D	✓	5	spy606
2013-02-16	Ajax File Manager Remote Code Execution Exploit	php	1947		R	D	⚠	25	tripleX
2013-02-16	A4tech Bloody2 Mouse Activation	tricks	1583		R	D	⚠	9	Alukard_X
2013-02-16	Dimofinf cms version 3.0.0 Sql Injection Vulnerability	php	1470		R	D	✓	30	spy606
2013-02-15	PHP-Nuke module (League 2.4) XSS Vulnerability	php	592		R	D	✓	2	GoLd_M
2013-02-15	PHP-Nuke Module Nukequiz <= 2.0.0 SQL Injection Vulnerability	php	714		R	D	✓	5	GoLd_M
2013-02-14	jibberbook Bypass Admin Vulnerability	php	1066		R	D	⚠	1780	CharafAnons
2013-02-13	Yahoo.com XSS Persistent + Cookie Exploit	tricks	1671		R	D	⚠	1920	paxx
2013-02-13	PostNuke Module phProfession <= 1.5 SQL Injection Vulnerability	php	364		R	D	⚠	5	GoLd_M
2013-02-13	Wordpress NextGEN Gallery 1.9.10 Arbitrary File Upload Exploit (win)	php	4468		R	D	⚠	20	bd0rk
2013-02-11	PayPal XSS + Cookie Stealer Exploit	tricks	5559		R	D	✓	2800	paxx
2013-02-11	phpBB highlight Arbitrary File Upload Vulnerability	php	1503		R	D	⚠	50	zhir
2013-02-10	Windows Service Pack 2 (explorer.exe) Memory Corruption	windows	1362		R	D	✓	30	The Black Devil..
2013-02-10	Wordpress Funny4You plugin 1.0 Local File Include Vulnerability	php	1589		R	D	✓	2	bd0rk
2013-02-09	Mozilla Firefox 18.0.2/Opera 12.12/Internet Explorer 9 Memory Corrupti..	multiple	1685		R	D	✓	25	The Black Devil..
2013-02-08	Wordpress privates themes (download.php) - Local File Inclusion	php	2106		R	D	✓	2	Zikou-16
2013-02-05	Facebook Privacy Vulnerability Create Private Messages from Anyone	tricks	19218		R	D	✓	700	buglab
2013-02-03	MS12-020 Remote Desk Top denial of service vulnerability (metasploit)	windows	2707		R	D	✓	250	The Black Devil..
2013-02-02	Apple Safari 6.0.2 (OS X) file:// Multiple Vulnerabilities	macOS	1329		R	D	✓	500	F1restorm_RST
2013-01-26	Wordpress plugins - slidedeck2 pro XSS/File Upload Vulnerability	php	2904		R	D	⚠	5	Zikou-16
2013-01-25	Microsoft Office 2003/2007/2010 Command Execution 0day	windows	12708		R	D	✓	5000	1337Day Team
2013-01-12	ColdFusion E-Commerce (SHOP) <= Local File Include Vulnerability	php	1811		R	D	✓	2	KedAns-Dz
2013-01-07	Joomla mega menu module File Upload Vulnerability metasploit	php	4693		R	D	✓	2	The Black Devil..
2013-01-01	Subrion CMS v2.3.x <= (FU/dDB) Multiple Vulnerabilities	php	639		R	D	✓	2	KedAns-Dz
2012-12-30	PrestaShop E-Commerce v1.5.x->1.5.2 Multiple Vulnerabilities	php	2984		R	D	✓	1	KedAns-Dz

Zero-Day-Exploit Market Example

[Detailed Information]

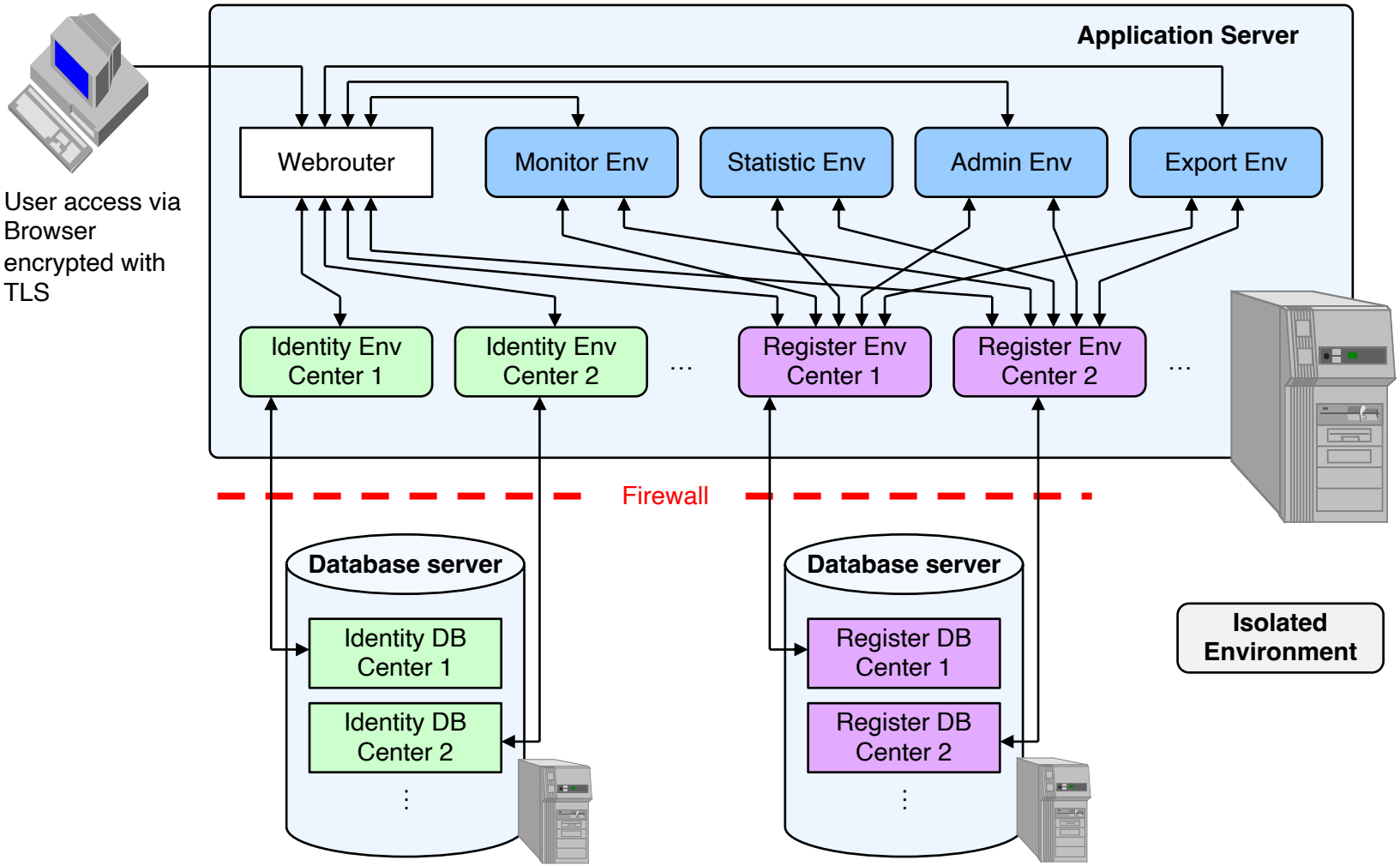
Full title:	Microsoft Office 2003/2007/2010 Command Execution 0day
Date add:	2013-01-25
Category:	remote exploits
Verified:	✓Verified
Risk:	
Platform:	windows
Views:	12708
Comments:	6
Price:	5000
Description:	Microsoft Office 2003/2007/2010 all service pack from a command execution vulnerability .
Youtube video:	
Rate up:	10 Rate up
Rate down:	3 Rate down
Warnings:	0

IDOMENEO approach for a Peripheral Artery Disease (PAD) Research Register

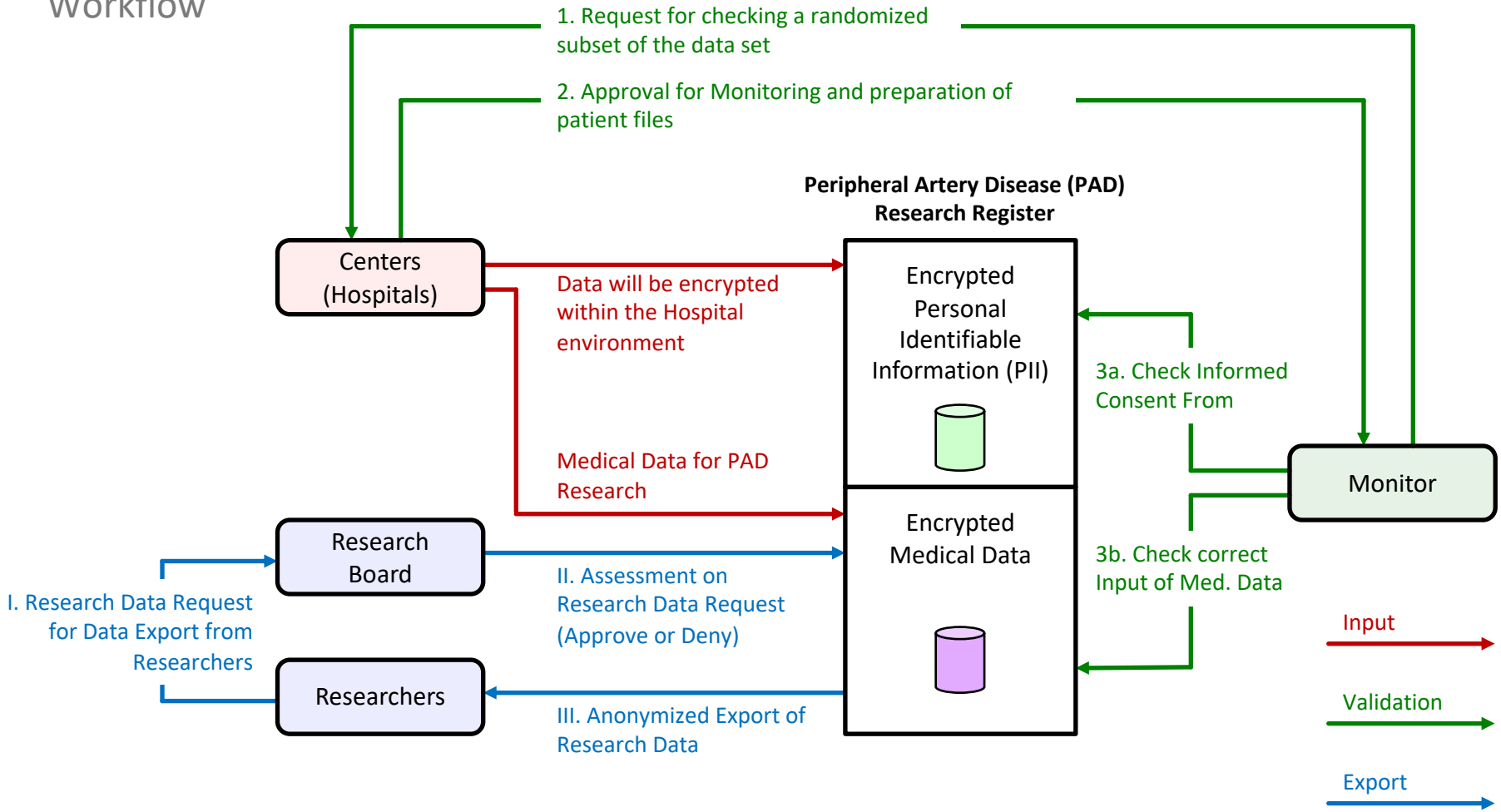


- Large number of Centers (Hospitals) allows an increased number of cases into the Research Data
- Informed Consent as a legal requirement for the inclusion of Study Participants (Patients)

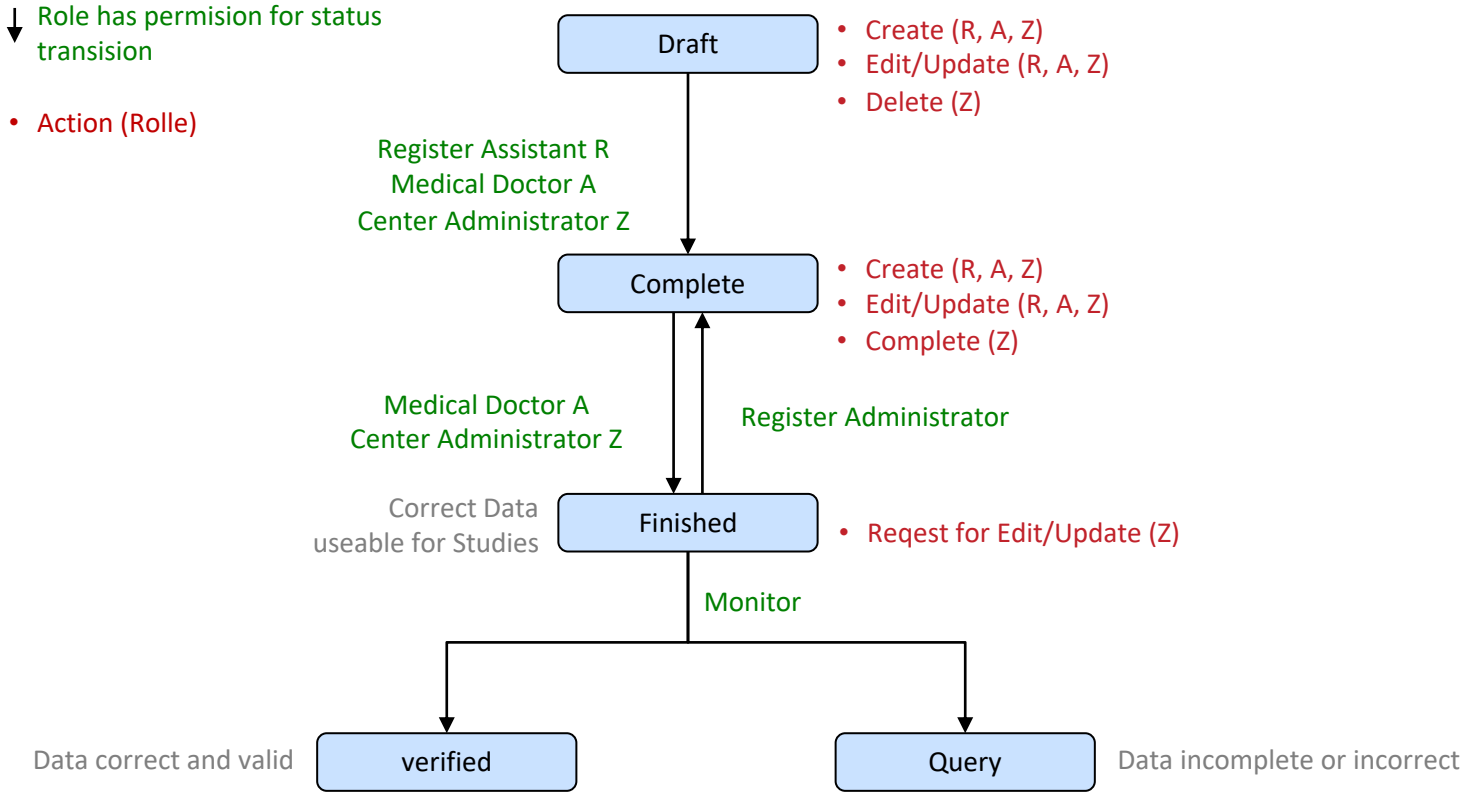
- Centers just use a modern Web Browser
- Centralized storage of Personal Data (strong symmetric encryption) and Medical data (also encrypted) within a Medical Center
- Operation and Maintenance (Updates, Malware protection) of Study Hardware in the hand of the Medical Research Register
- Centers own a symmetric key, extracted via Password Based Key Derivation Function (PBKDF)


Application Server
Webrouter
Monitor Env
Statistic Env
Admin Env
Export Env
User access via Browser encrypted with TLS
Identity Env Center 1
Identity Env Center 2
...
Register Env Center 1
Register Env Center 2
...
Firewall
Database server
Identity DB Center 1
Identity DB Center 2
Database server
Register DB Center 1
Register DB Center 2
Isolated Environment

Workflow


1. Request for checking a randomized subset of the data set
2. Approval for Monitoring and preparation of patient files
Peripheral Artery Disease (PAD) Research Register
Centers (Hospitals)
Data will be encrypted within the Hospital environment
Encrypted Personal Identifiable Information (PII)
3a. Check Informed Consent From
Monitor
Medical Data for PAD Research
Encrypted Medical Data
3b. Check correct Input of Med. Data
Research Board
I. Research Data Request for Data Export from Researchers
II. Assessment on Research Data Request (Approve or Deny)
Researchers
III. Anonymized Export of Research Data
Input
Validation
Export

Status / Livecycle of a data entry



Selected Challenges (Operation)

- How to ensure long-term security/encryption (> 10 years)?
  - Use symmetric cryptographic systems only on core system
  - Limitation: Browser support, TLS uses public key encryption
  - Encrypted layer on top of TLS (Layer 7 encryption)
  - Modular encryption functionalities allow easily changing to state-of-the-art algorithms

- How to protect application server from curious admins during maintenance?
  - Technical staff (server admins) is not allowed to learn about PII and medical data stored
  - Limitation: Admin has root access to server
  - Monitored access (audit trails) during maintenance and double encrypted data base

- How to ensure safe environment on the medical center (client) side?
  - Security depends on secrecy of password (and security code) – only known by medical center
  - Limitation: No malware or (insider) attacker present on local machine?
  - Trust model: Medical centers are aware of secure configuration and operation of hardware

Select appropriate key lengths and algorithms?

»SOG-IS (Senior Officials Group Information Systems Security) Crypto Evaluation Scheme Agreed Cryptographic Mechanisms« gives official information

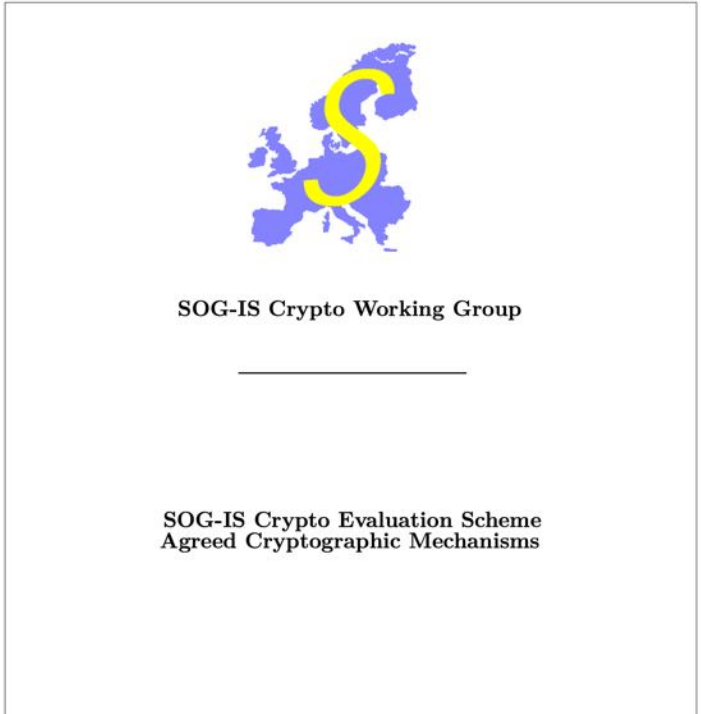
https://www.sogis.eu/documents/cc/crypto/SOGIS-Agreed-Cryptographic-Mechanisms-1.2.pdf

Agreed Block Ciphers.

Primitive	Parameters' sizes	R/L	Notes
AES [FIPS197, ISO18033-3]	k = 128 bits	R	
	k = 192 bits	R	
	k = 256 bits	R	

Agreed Hash Functions

Primitive	Parameters' sizes (hash length h)	R/L	Notes
SHA-2 [FIPS180-4, ISO10118-3]	h = 256 bits (SHA-256)	R	
	h = 384 bits (SHA-384)	R	
	h = 512 bits (SHA-512)	R	
	h = 256 (SHA-512/h)	R	
SHA-3 [FIPS202]	h = 256 bits	R	
	h = 384 bits	R	
	h = 512 bits	R	



SOG-IS Crypto Working Group

SOG-IS Crypto Evaluation Scheme
Agreed Cryptographic Mechanisms

Brute-force, exhaustive search

- Attacks via Super Computers and (in the future) Quantum Computers
  - Complexity theoretic systems only
- Protection against Super Computers
  - Use appropriate key lengths
- Protection against Quantum Computers
  - Symmetric systems: Double key lengths to >= 256 Bit
  - Asymmetric systems: Hope in Post-Quantum Cryprography

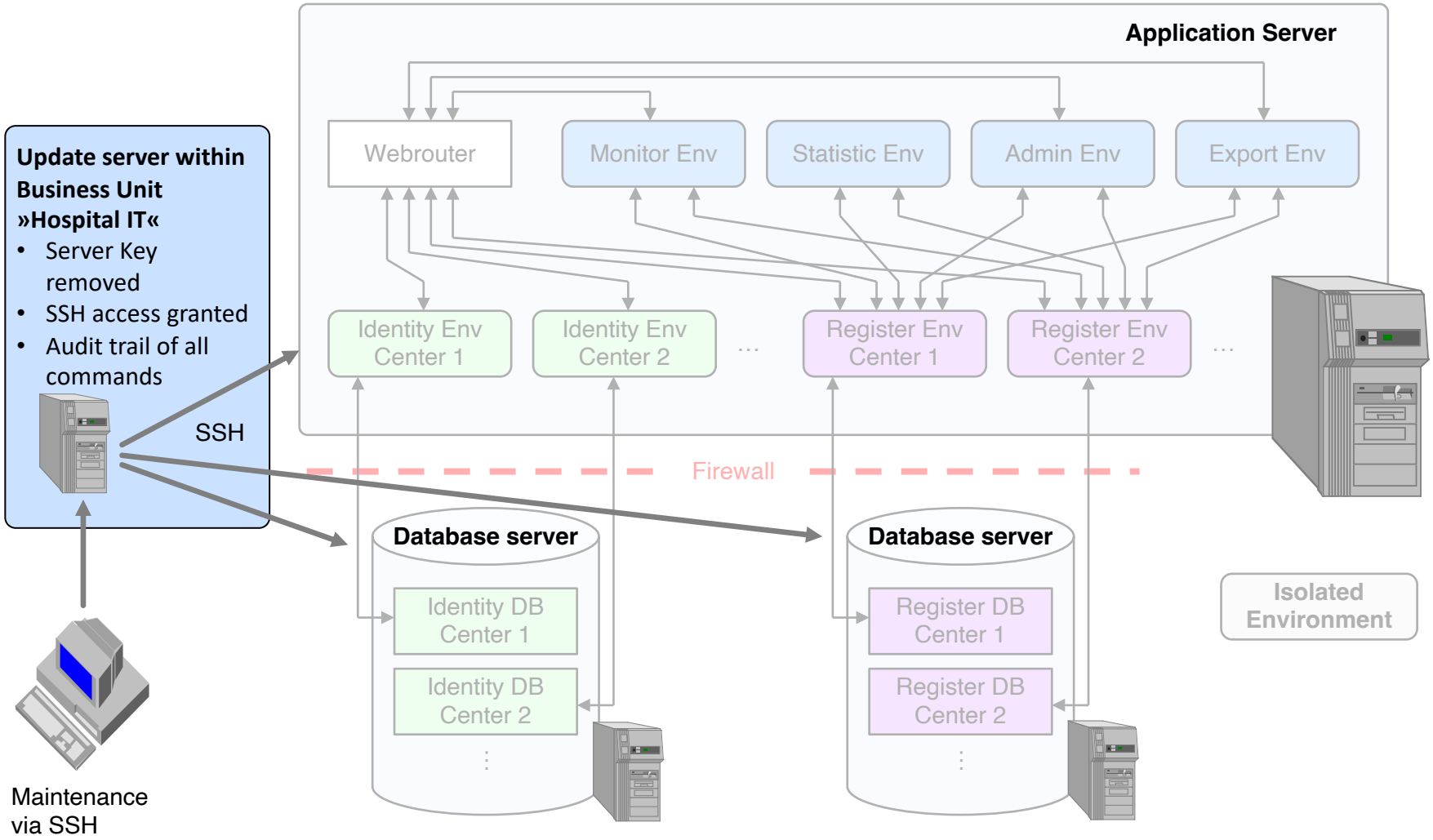
	Key lengths	Complexity		According to: Bernstein, Buchmann, Dahmen: Post Quantum Cryptography. Springer, 2009
		Super Computer	Quantum Computer	
Symm.	128 Bit	2^{127}	2^{64}	Grover, 1996
	256 Bit	2^{255}	2^{128}	
Asymm.	1024 Bit	$\approx 2^{90}$	$\approx 2^{25}$	Shor, 1994
	2048 Bit	$\approx 2^{117}$	$\approx 2^{28}$	

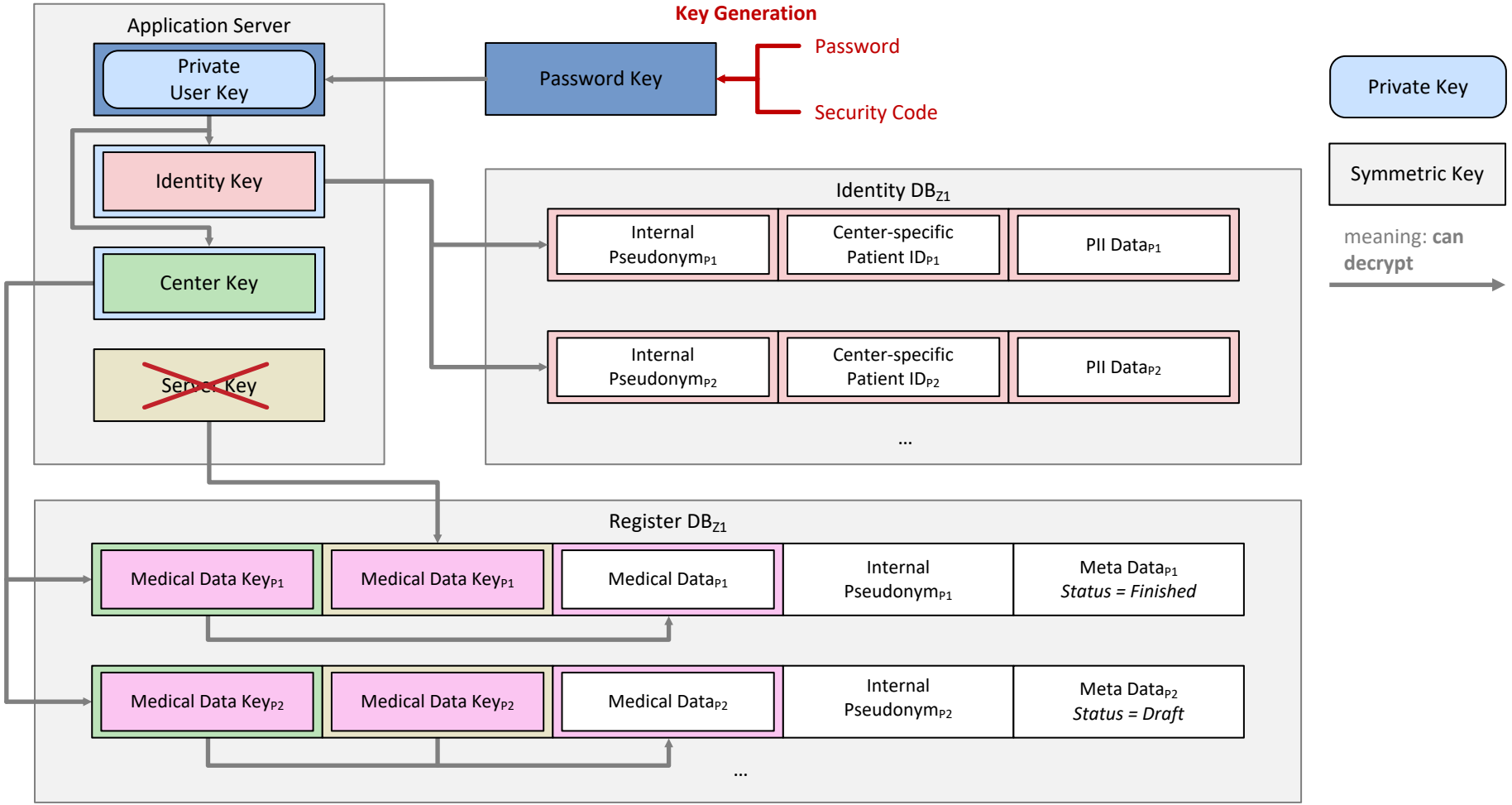
Selected Challenges (Operation)

- How to ensure long-term security/encryption (>= 10 years)?
  - Use symmetric cryptographic systems only on core system
  - Limitation: Browser support, TLS uses public key encryption
  - Encrypted layer on top of TLS (Layer 7 encryption)
  - Modular encryption functionalities allow easily changing to state-of-the-art algorithms

- How to protect application server from curious admins during maintenance?
  - Technical staff (server admins) is not allowed to learn about PII and medical data stored
  - Limitation: Admin has root access to server
  - Monitored access (audit trails) during maintenance and double encrypted data base

- How to ensure safe environment on the medical center (client) side?
  - Security depends on secrecy of password (and security code) – only known by medical center
  - Limitation: No malware or (insider) attacker present on local machine?
  - Trust model: Medical centers are aware of secure configuration and operation of hardware


Application Server
Webrouter
Monitor Env
Statistic Env
Admin Env
Export Env
Identity Env Center 1
Identity Env Center 2
Register Env Center 1
Register Env Center 2
Update server within Business Unit »Hospital IT«
• Server Key removed
• SSH access granted
• Audit trail of all commands
SSH
Firewall
Database server
Identity DB Center 1
Identity DB Center 2
Database server
Register DB Center 1
Register DB Center 2
Isolated Environment
Maintenance via SSH

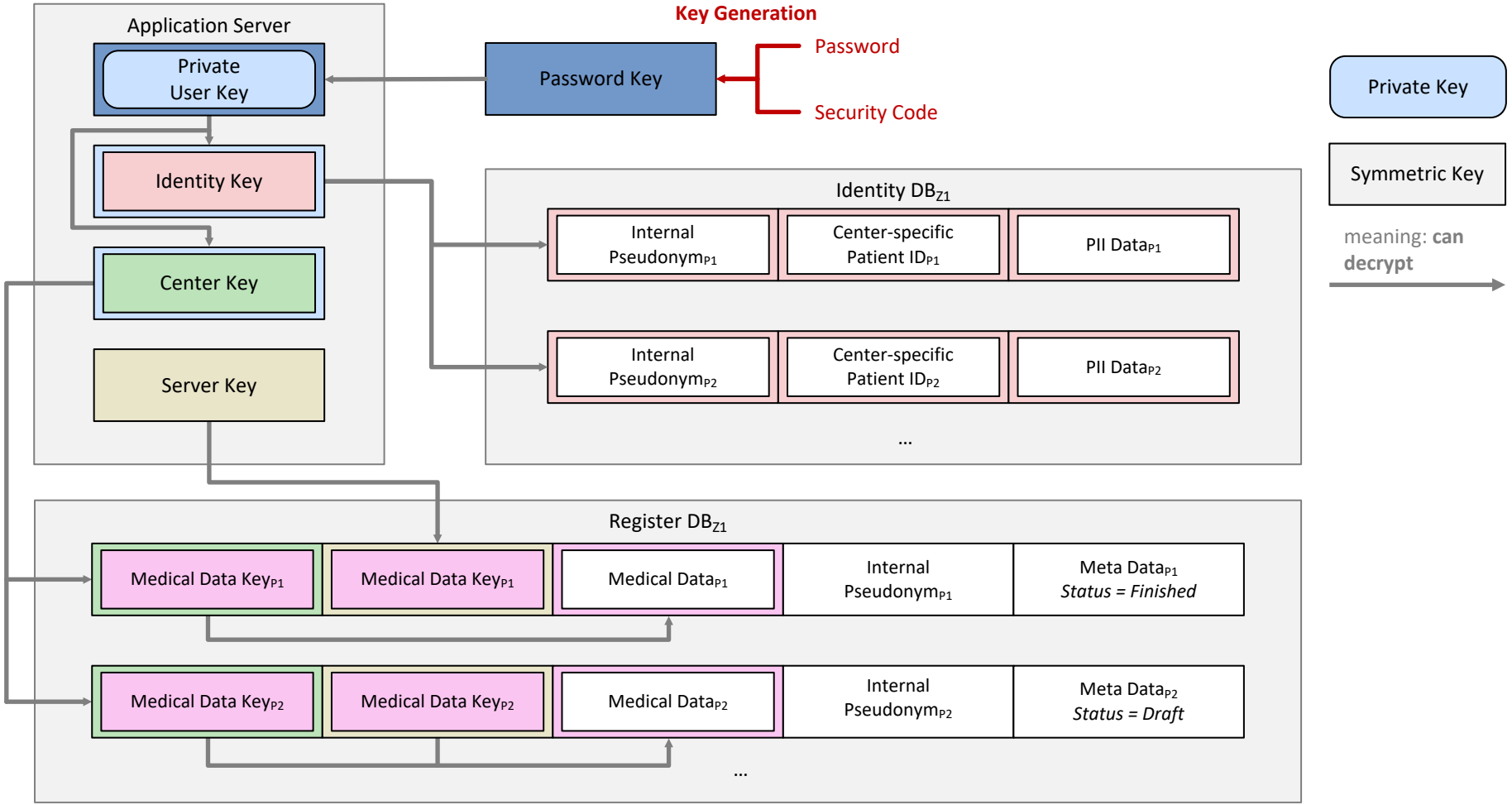

Application Server
Private User Key
Identity Key
Center Key
Service Key
Key Generation
Password Key
Password
Security Code
Identity DB$_{Z1}$
Internal Pseudonym$_{P1}$
Center-specific Patient ID$_{P1}$
PII Data$_{P1}$
Internal Pseudonym$_{P2}$
Center-specific Patient ID$_{P2}$
PII Data$_{P2}$
...
Register DB$_{Z1}$
Medical Data Key$_{P1}$
Medical Data Key$_{P1}$
Medical Data$_{P1}$
Internal Pseudonym$_{P1}$
Meta Data$_{P1}$
Status = Finished
Medical Data Key$_{P2}$
Medical Data Key$_{P2}$
Medical Data$_{P2}$
Internal Pseudonym$_{P2}$
Meta Data$_{P2}$
Status = Draft
...
Private Key
Symmetric Key
meaning: **can decrypt**

Selected Challenges (Operation)

- How to ensure long-term security/encryption (>= 10 years)?
  - Use symmetric cryptographic systems only on core system
  - Limitation: Browser support, TLS uses public key encryption
  - Encrypted layer on top of TLS (Layer 7 encryption)
  - Modular encryption functionalities allow easily changing to state-of-the-art algorithms

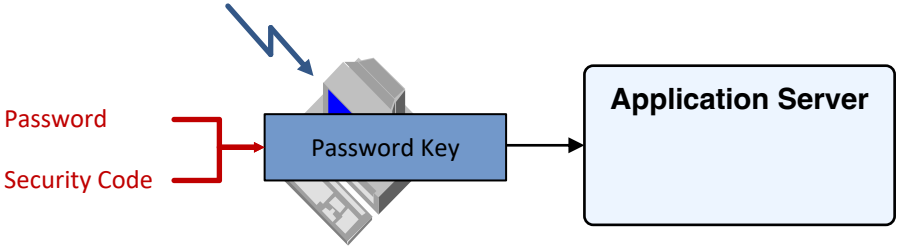
- How to protect application server from curious admins during maintenance?
  - Technical staff (server admins) is not allowed to learn about PII and medical data stored
  - Limitation: Admin has root access to server
  - Monitored access (audit trails) during maintenance and double encrypted data base

- How to ensure safe environment on the medical center (client) side?
  - Security depends on secrecy of password (and security code) – only known by medical center
  - Limitation: No malware or (insider) attacker present on local machine?
  - Trust model: Medical centers are aware of secure configuration and operation of hardware


Application Server
Private User Key
Identity Key
Center Key
Server Key
Key Generation
Password Key
Password
Security Code
Identity DB$_{Z1}$
Internal Pseudonym$_{P1}$
Center-specific Patient ID$_{P1}$
PII Data$_{P1}$
Internal Pseudonym$_{P2}$
Center-specific Patient ID$_{P2}$
PII Data$_{P2}$
...
Register DB$_{Z1}$
Medical Data Key$_{P1}$
Medical Data Key$_{P1}$
Medical Data$_{P1}$
Internal Pseudonym$_{P1}$
Meta Data$_{P1}$
Status = Finished
Medical Data Key$_{P2}$
Medical Data Key$_{P2}$
Medical Data$_{P2}$
Internal Pseudonym$_{P2}$
Meta Data$_{P2}$
Status = Draft
...
Private Key
Symmetric Key
meaning: **can decrypt**

How to ensure safe environment on the medical center (client) side?

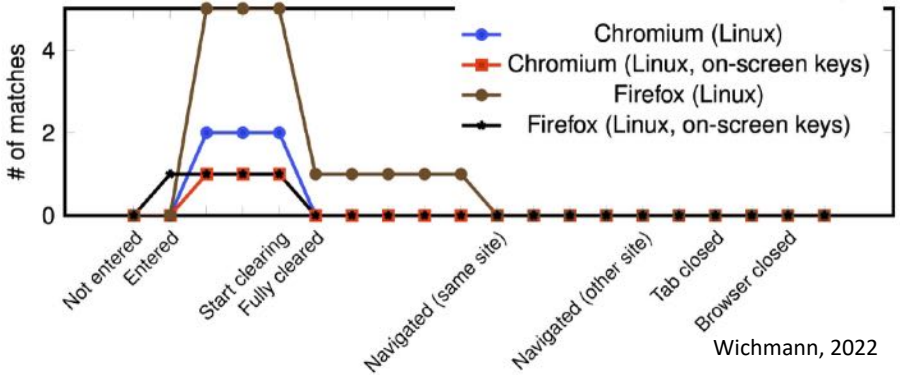
- No malware or (insider) attacker present on local machine?
- Medical centers are aware of secure configuration and operation of hardware?



User access via Browser encrypted with TLS

SecPassInput:
Content stored in UInt8Array

Type PW and delete, not send

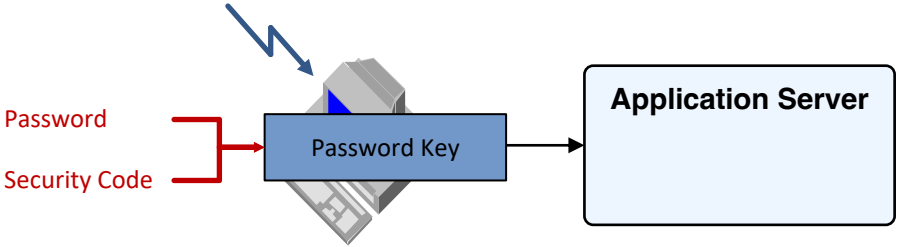


Wichmann, 2022

- Question: Is it possible to reconstruct Password Key from local memory via targeted attack?
  - Malicious application on Client frequently
    - Dump memory
    - Analyzing swap files
    - Cold boot attack (memory remanation)

How to ensure safe environment on the medical center (client) side?

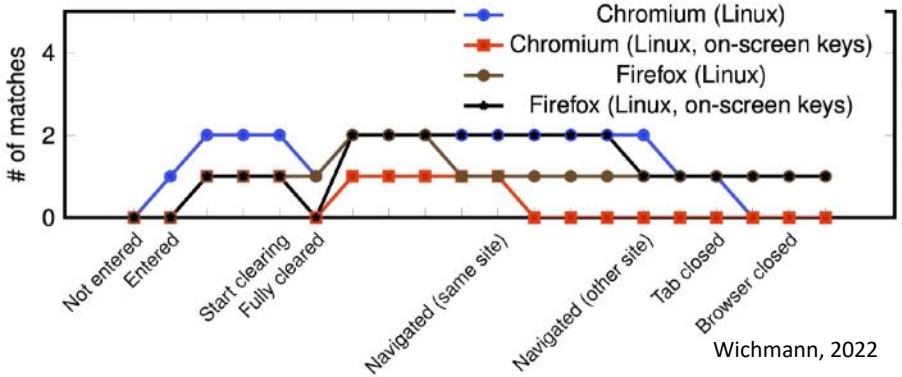
- No malware or (insider) attacker present on local machine?
- Medical centers are aware of secure configuration and operation of hardware?



User access via Browser encrypted with TLS

SecPassInput:
Content stored in UInt8Array

Type PW, send, and programmatically clear memory



Wichmann, 2022

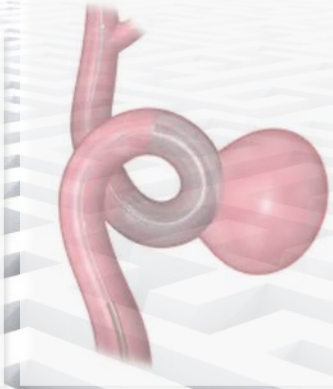
- Question: Is it possible to reconstruct Password Key from local memory via targeted attack?
  - Malicious application on Client frequently
    - Dump memory
    - Analyzing swap files
    - Cold boot attack (memory remanation)

Privacy-Preserving Design and Operation of Medical (Research) Registers

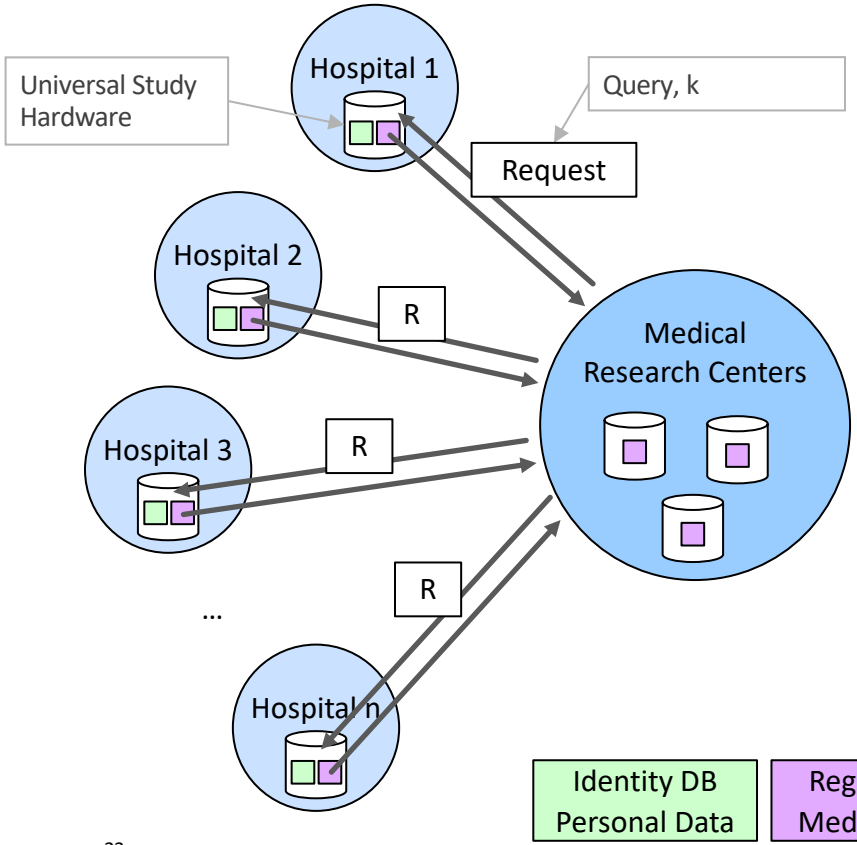
- Introduction: Classical approach to the realization of Medical Research Registers
- IDOMENEO approach for a Peripheral Artery Disease (PAD) Research Register
  - Architecture and Workflow
  - Selected Challenges
- PANDA approach for device studies (stents, balloons) in neuroradiology (strokes and aneurysms)
  - Decentralized storage, ready for (federated) machine learning approaches
  - Prepare for upcoming Research Data Law (data usage without informed consent)
- Conclusions and future research directions

Gemeinsamer Bundesausschuss

Bundesministerium für Wirtschaft und Energie



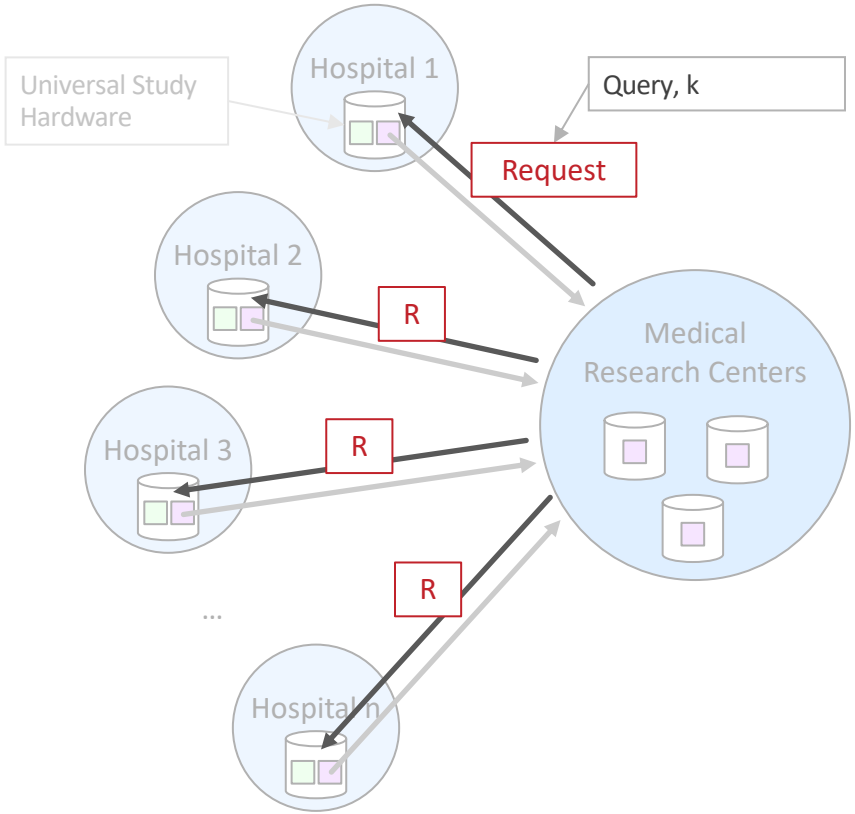
PANDA approach for device studies (stents, balloons) in neuroradiology



- Large number of Centers (Hospitals) allows an increased number of cases into the Research Data
- Informed Consent as a legal requirement for the inclusion of Study Participants (Patients)

- Data records remain in Centers
- Approach to distributed gathering of research data needed for a concrete research question
- Adaptive k-anonymous response from centers
- Limited number of (similar) requests
- Makes use of privacy-respecting federated learning and secure multi-party computation (SMPC)

Agent k creator

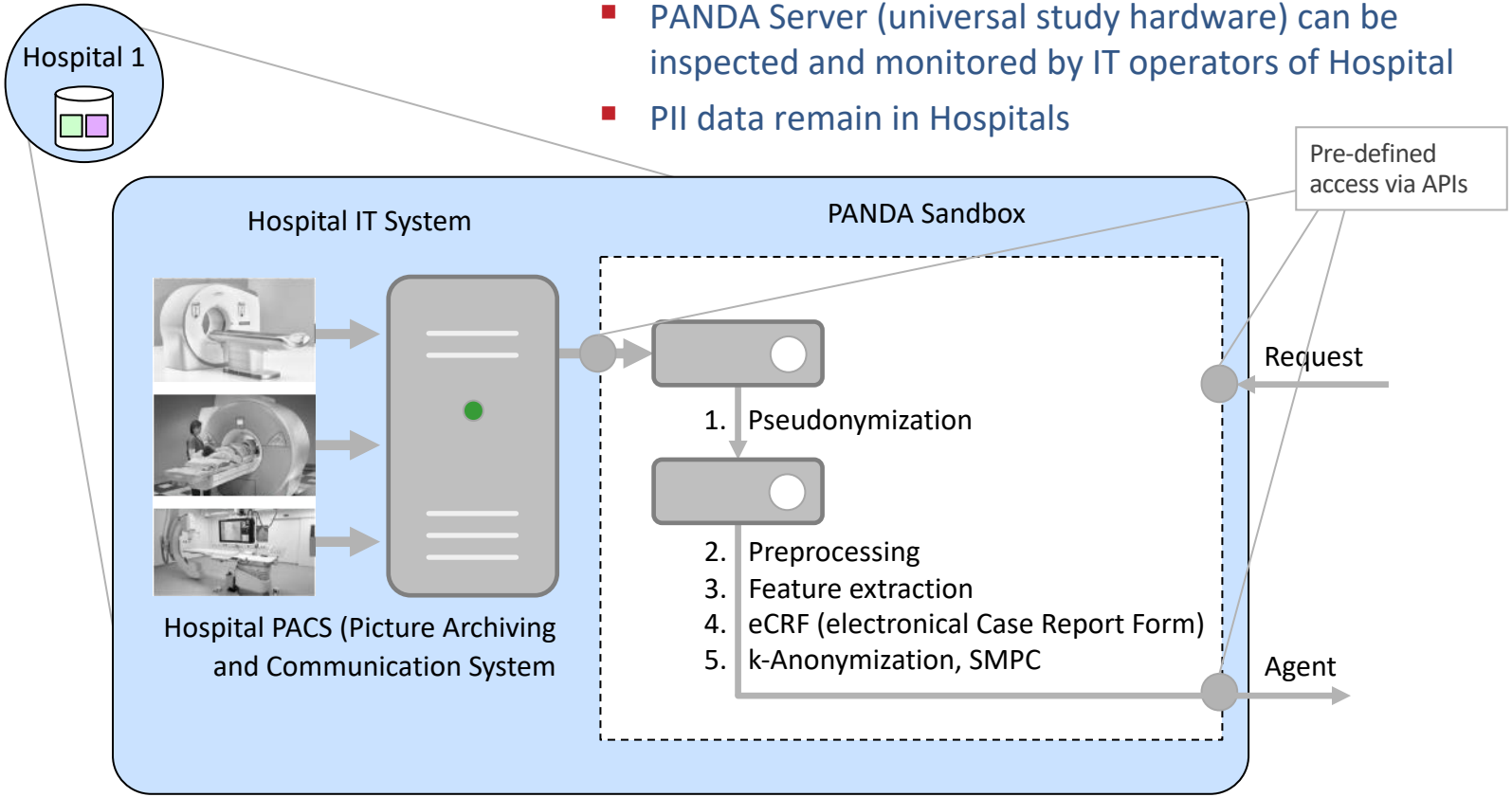


- Central definition of requested variables, their range and their granularity
- Type of query (also repetitive query) with pre-defined k or patient number

	Range	Granularity
Device	Solitaire Stent	4 x 40 mm only
Age	0 18 70	10
Symptoms (NIHSS-Score)	0 8 20	2
Infarct volume (ASPECT-Score)	7 10	2
◉ K	3	○ Patients 297

PANDA hardware architecture in Hospitals

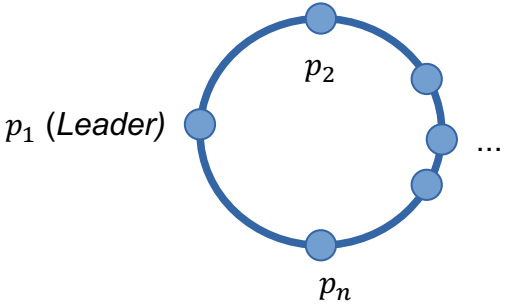
- PANDA Server (universal study hardware) can be inspected and monitored by IT operators of Hospital
- PII data remain in Hospitals



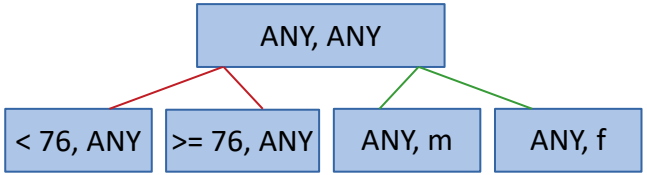
Employing SMPC for Decentralized Anonymisation

- Preliminary work [Mohammed et al., 2010] is very efficient, however round-based and insecure
- Proposal of an extended algorithm for Decentralized Anonymization

- Top-down approach by [Mohammed et al., 2010]
  - Begin with fully generalized data (e.g., AGE = ANY, SEX = ANY, ...)
  - One specialization per round (e.g., split ages < 76 and >= 76) based on count statistics gathered via secure sum protocol
  - Specialize until each possible specialization would violate k-anonymity
- Count statistics
  - How many records are contained in each equivalence class?
  - For each equivalence class: How many records would be in sub-classes if this class was specialized?



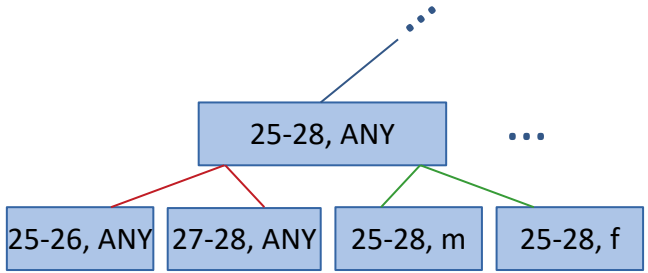

```
{'age.ANY|sex.ANY|': (45220, {
age: {'0:75': 36539, '76:120': 8681},
sex: {'Female': 14695, 'Male': 30525}
})}
```



Leaking data in the basic protocol

[Mohammed et al., 2010]

- Final count statistics leak data to the leading party
- This leaks more information than necessary which can **break the k-anonymity guarantees** completely, e.g., when combined with a background knowledge attack
- Secure sum protocol insecure, when parties p_{i-1} and p_{i+1} collaborate



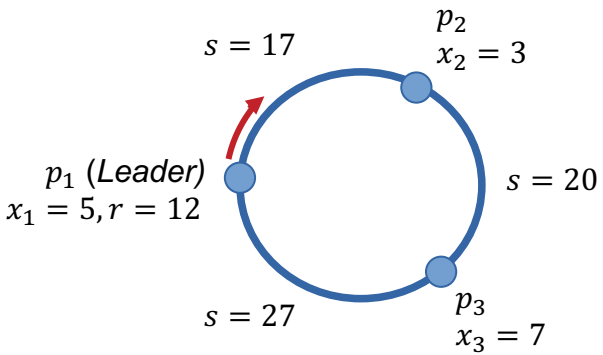

```
{'age.25-28|sex.ANY|': (6, {
age: {'25-26': 2, '27-28': 4},
sex: {'f': 1, 'm': 5}
})}
```

Sex	Age	Cancer Diagnosis
ANY	25-28	Breast
ANY	25-28	Lungs
ANY	25-28	Brain
ANY	25-28	Testicle
ANY	25-28	Lungs
ANY	25-28	Skin

Alternative Approach

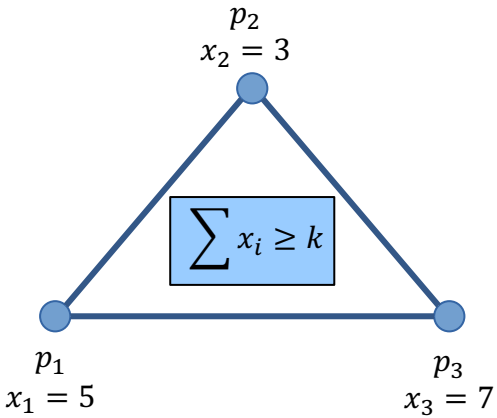
Secure Sum Protocol

- Each party has an input value x_i
- We want to learn the sum of all inputs without disclosing single inputs



Secure Multiparty Computation

- Each party has an input value x_i
- We want to learn **if** the sum of all inputs **is larger than k** without disclosing single inputs **or the sum itself**



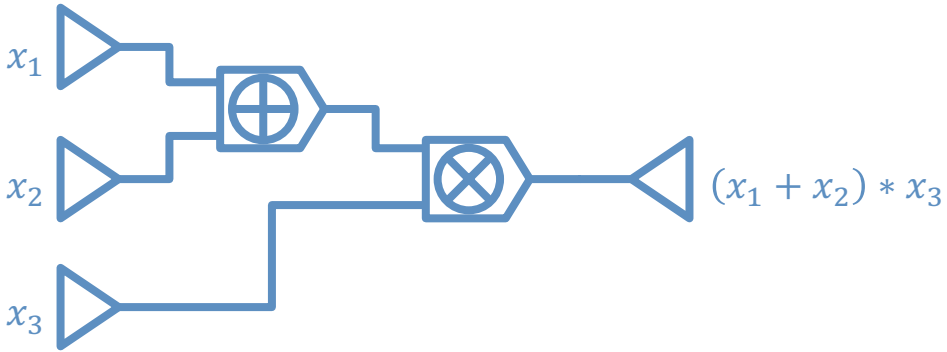
Arithmetic Circuits

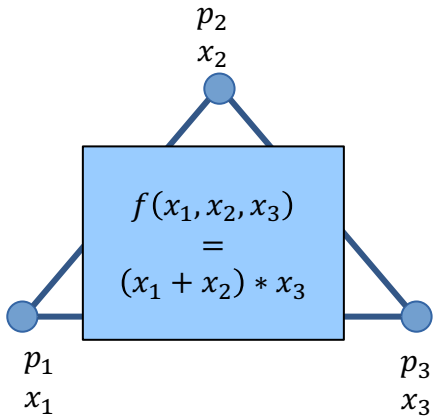
- Arithmetic circuits (comprised of multiplication and addition gates) are built over a finite field $\mathbb{Z}_p$ with $p > n$ for n being the number of parties.
- Arithmetic circuits are Turing complete, so any function can be represented via these circuits.

Addition gate

Multiplication gate

Example



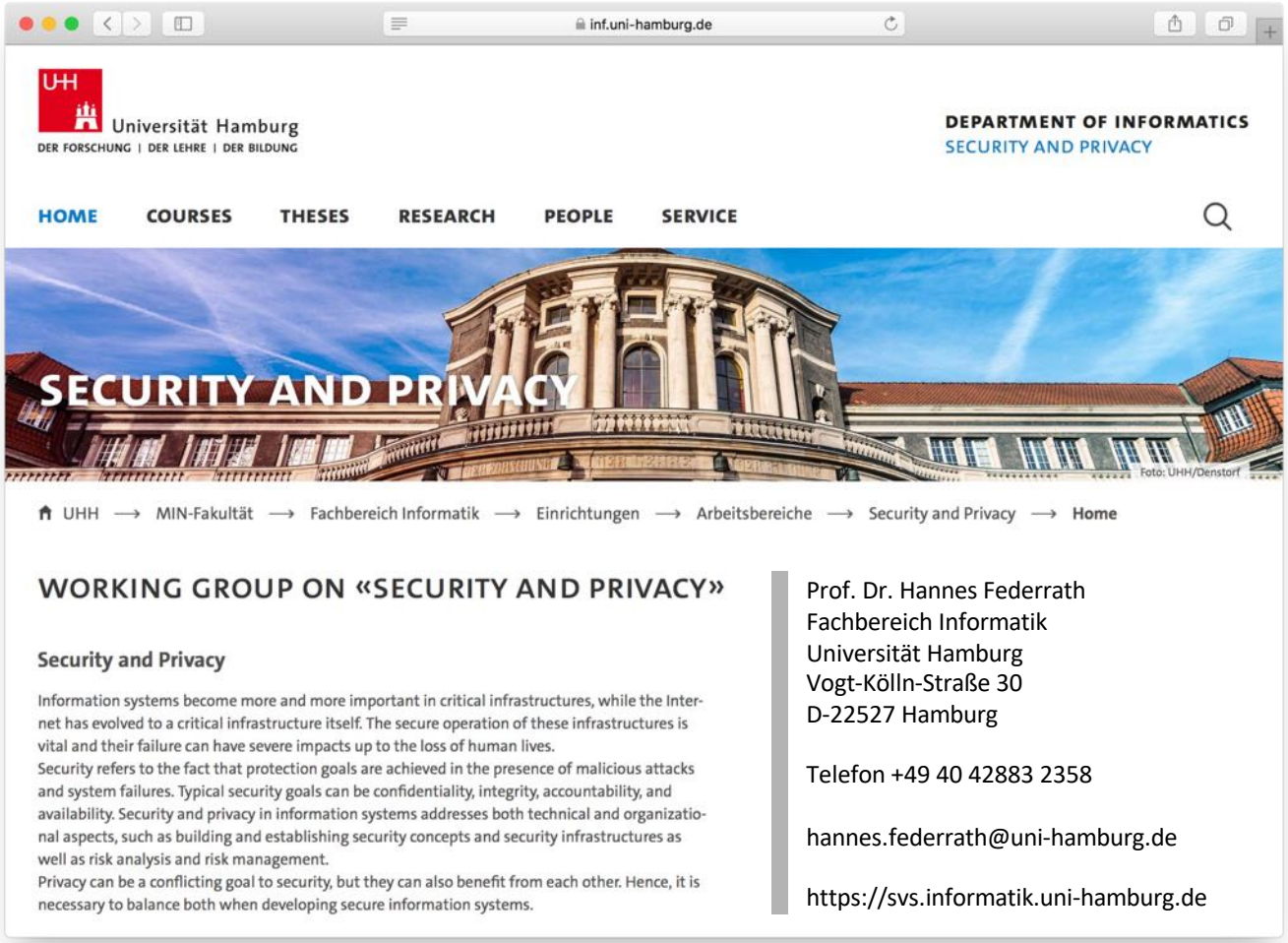


Conclusions and future research directions

- Using general SMPC frameworks is possible, but can have a large overhead in communication and computation
  - might be acceptable in the concrete scenario since this protocol has to be performed exactly once for a data set to be published afterwards

- Overall: very high complexity of Medical Research Registers due to legal requirements
  - Ethical approval
  - IT security concept, Privacy concept, Role concept, Risk assessment
  - Privacy Impact Analysis (PIA) (within KI-SIGS, BMWI)

- New research project (BMBF, 2023-2026): AnoMed
  - Benchmarking of Anonymization standards for Medical Data



With thanks to:
Christian Behrendt
Daniel Demmler
Jens Fiehler
Tobias Müller
Tom Petersen
Henning Pridöhl
Pascal Wichmann
…