Leiden University Medical Center

ProMISe

Privacy by design and (big or not-so-big) clinical research data (management)

Safeguarding privacy while maximizing scientific benefits: a biostatistician's approach to good data management

Ronald Brand

Dep of Medical Statistics, section Advanced Data Management

Leiden University Medical Center

R.BRAND@LUMC.NL

Care

Research

Privacy

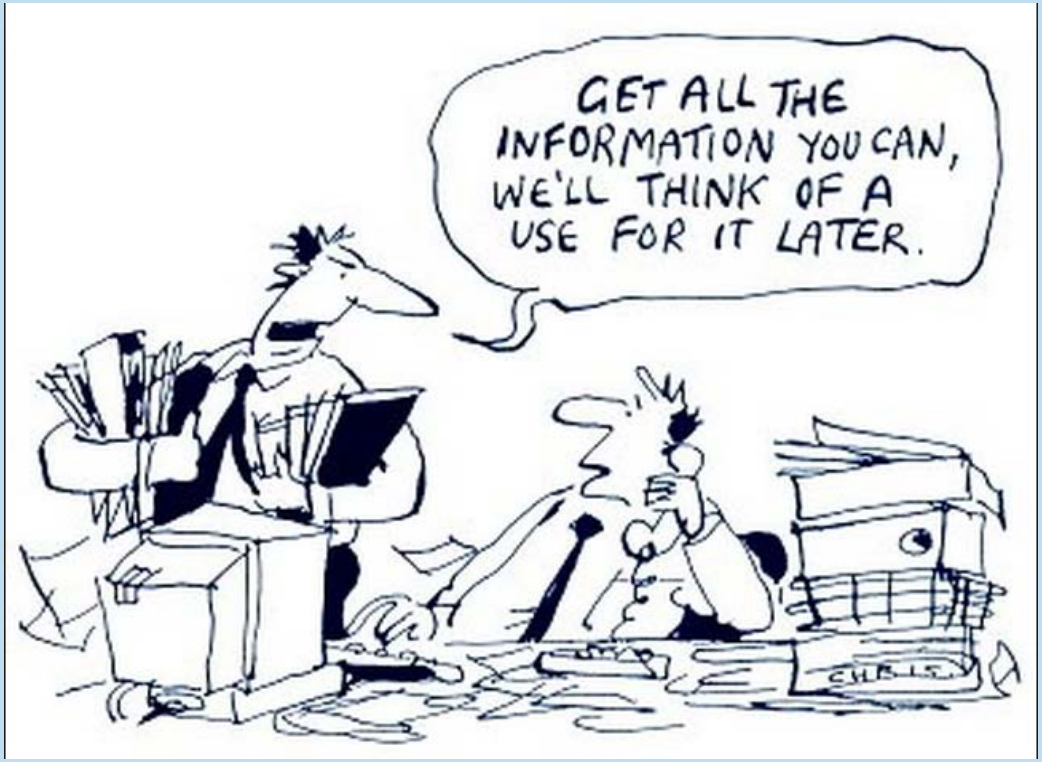
Traceability

the god of beginnings, gates, transitions, time, doorways, passages, and endings

What *is* (big) data management?

- *Section Medical Statistics*
  - Statistical consultation for LUMC + others
    - Clinical trials
      - Design
      - Analysis
      - Data Safety and Monitoring Board
      - Medical Ethical Committee
  - Teaching
  - Research
- *Section Advanced Data Management*
  - Provide secure, advanced, cost-effective, web based data management infrastructures for clinical research
  - Make sure design facilitates the intended analyses as well as the intended users, maximizing privacy protection

NATURE AND PURPOSE OF DATA COLLECTION IN CLINICAL RESEARCH



Data collection types & follow-up

- **Observational (cohort) data**
  - Just “observe”; do not interfere with treatment or impose different behavior
- **Experimental designs**
  - Modify treatment/behavior according to protocol
- **Quality registers**
  - Use care data for improvement of care/clinical research

The notion of follow-up in outcome measurement

The very notion of “development of health and illness” requires the researcher to follow the patient through time and space. This will inherently invade his or her privacy so protect the process by all means.

Design of studies and type of privacy issues

- clinical trials
- cohort studies
- transition of data from Care to Research
- quality registers
- rare diseases
- mixtures: registries to support bot
- ultra-sensitive registries

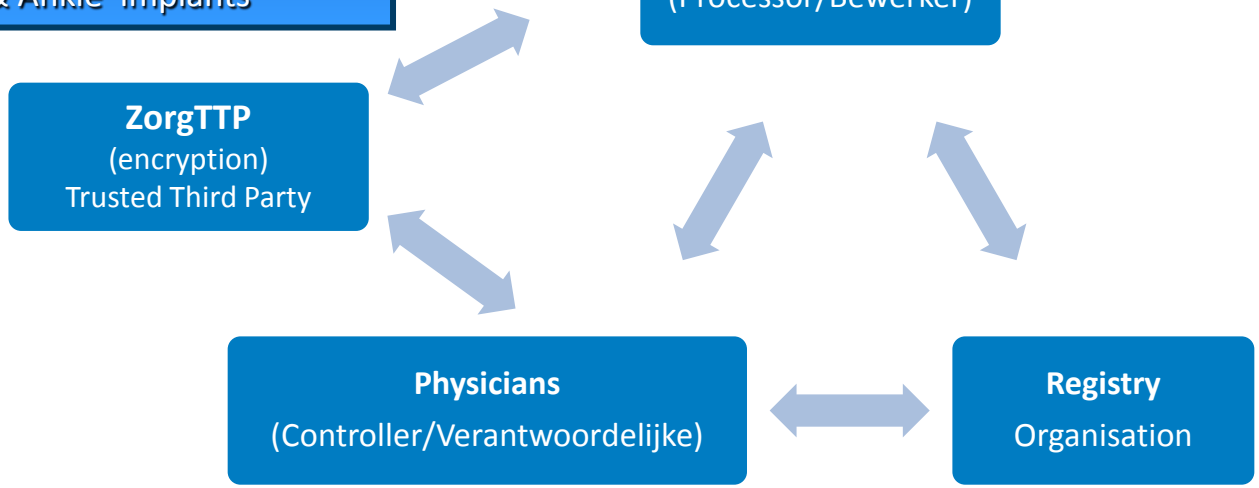
Protection by …

- Account (role) management
- encryption
- transparency => trust
- Principle of necessity, proportionality and subsidiarity

Quality Registers: compare devices

>430.000 patients, 290000 hips, 290000 knees

LROI: National Registry of all Hip & Knee & Wrist & Shoulder & Ankle implants



Privacy aspects: care data; comparison of devices on outcome; sensitive data for patients, hospitals, industries

Solution: encrypted identities for patients; contracts between all hospitals and data base host (LUMC) as well as between LUMC and Foundation as well as participation of physicians in Foundation; privacy committee; scientific committee; informed opt-out

Trauma Registry

landelijk netwerk acute zorg

National and regional registries of all accidents in the Netherlands

>750.000 incidents, fully classified according to AIS score

DIAGNOSE GEGEVENS		
Diagnose		
Diagnose regio	63	Abdom
Diagnose categorie	2991	Spleen
AIS-code		
Omschrijving		

5442102	[Abdom] {Spleen} contusion (hematoma) NFS
5442122	[Abdom] {Spleen} contusion (hematoma) minor (superficial; subcapsular; < = 50% surface area; OIS Gr
5442143	[Abdom] {Spleen} contusion (hematoma) major (subcapsular > 50% surface area or expanding) OIS Gr
5442202	[Abdom] {Spleen} laceration NFS
5442222	[Abdom] {Spleen} laceration minor (superficial) OIS Grade I or II
5442243	[Abdom] {Spleen} laceration moderate (no hilar or segmental parenchymal disruption or destruction) OI
5442264	[Abdom] {Spleen} laceration major (involving segmental parenchymal disruption or destruction with no h IV
5442285	[Abdom] {Spleen} laceration massive (with hilar disruption) OIS Grade V
5442403	[Abdom] {Spleen} rupture (`fracture`) NFS
5442992	[Abdom] {Spleen} NFS

Privacy aspects: required by law; data from patient care; Goal: science&quality

Solution: fully encrypted; covered by contracts

Rare diseases, mixture registries, ultra-sensitive

rare diseases

- May easily lead to identifiability, hence anonymity is a myth

mixtures: registries to support both quality improvement and science

- Not trivial: the use of the *same* data for *different* purposes
  - Quality improvement by analyzing your own data: that is even mandatory!
  - Quality improvement by comparing your data to others: either informed consent or *informed* opt-out or anonymization needed
  - Scientific Research: anonymization feasible (and thus mandatory)

Some aspects of data collections

- Quality of data
- Missing data
- Follow-up
- Selection bias
  - Informed consent
  - Informed opt-out

  Case law / jurisprudence?

"Before I write my name on the board, I'll need to know how you're planning to use that data."

TENSION

Need to increase scientific knowledge versus need to maintain privacy for patient, physician and institute contributing to that knowledge

Our legal system: what do I have to pay attention to?

ProMISe

- WBP Wet Bescherming Persoonsgegevens (Personal Data Protection Act)
- BIG Wet op de Beroepen in de Individuele Gezondheidszorg (Individual Healthcare Professions Act)
- WGBO Wet op de Geneeskundige Behandelingsovereenkomst (Medical Treatment Contracts Act)
- WPR Wet Persoonsregistraties (Personal Data Files Act)
- CBP College Bescherming Persoonsgegevens; now: Autoriteit Persoonsgegevens (Data Protection Authority)

Essential starting points:

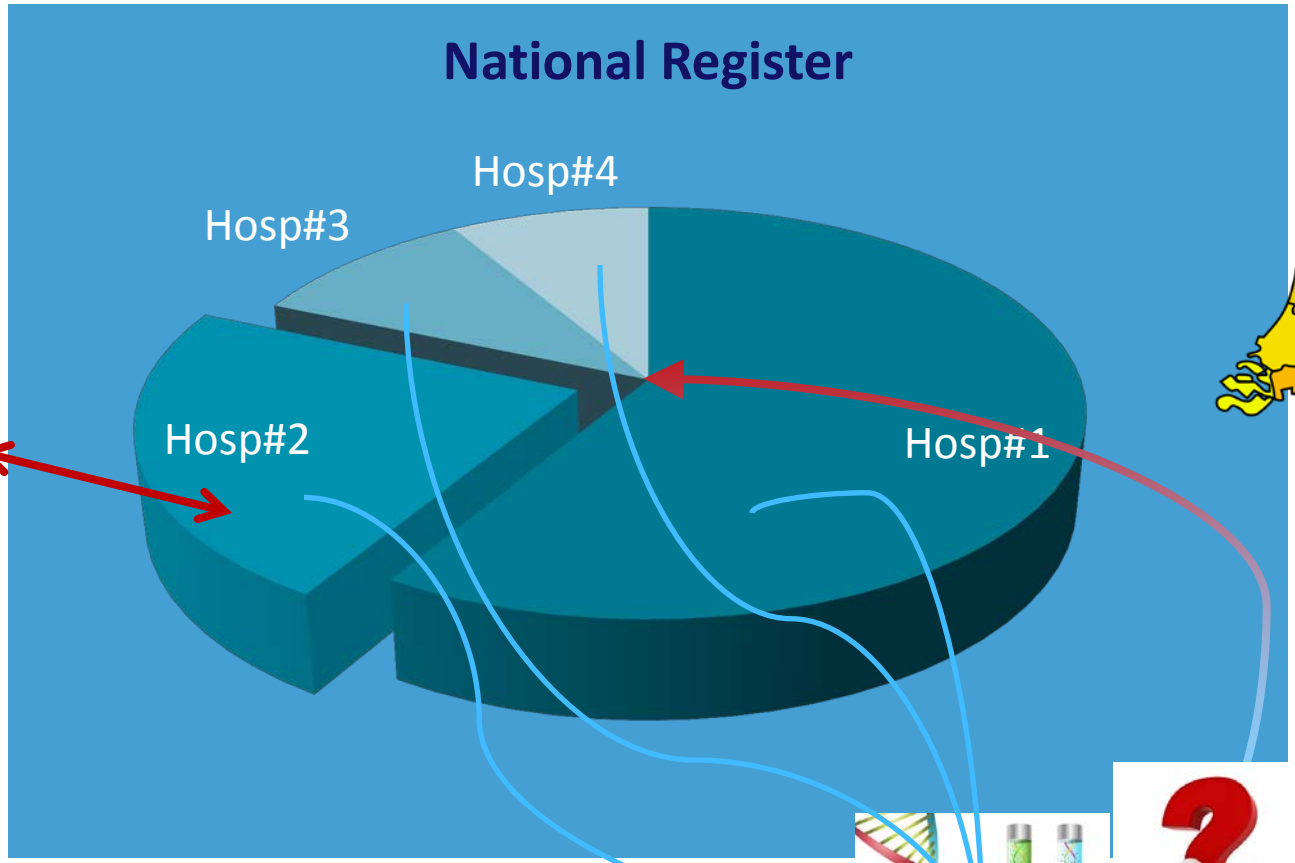
- Medical files should be accessible only by those who provide care
- Research data bases should not contain direct person identifiers unless explicitly allowed by the law and made inaccessible to those without a “need to know”
- Never store in a data base or file what you do not really need to fulfill the goal of your research project

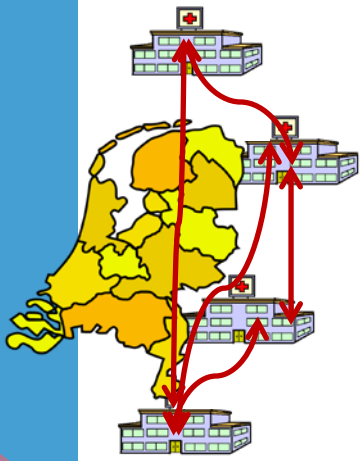
Legal framework of Quality of Care comparisons

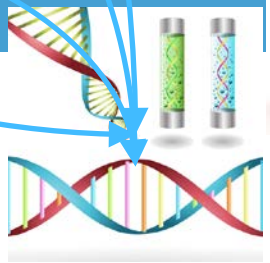
Interesting situation from a data protection (legal) point of view

- Data are provided from the Care Domain with the purpose of Quality enhancement
  - If goal is comparison of one's own data to the (adjusted) average, it is called "care" and the legal system surrounding data protection in health care applies
  - If goal is to enhance quality of care nationwide, through comparison of multiple centers, the storage of data is still from a "Care perspective" but the use of data is governed by the usual "Data Protection Act" but still in the framework of Care
  - If goal is to enhance effectiveness of care and improvement through scientific interpretation, the whole framework of "Clinical Research" applies
  - **Storage** may be subject to one legal safety net, the **use** and **access** might be governed by another legal system

Legal framework of Quality of Care comparisons





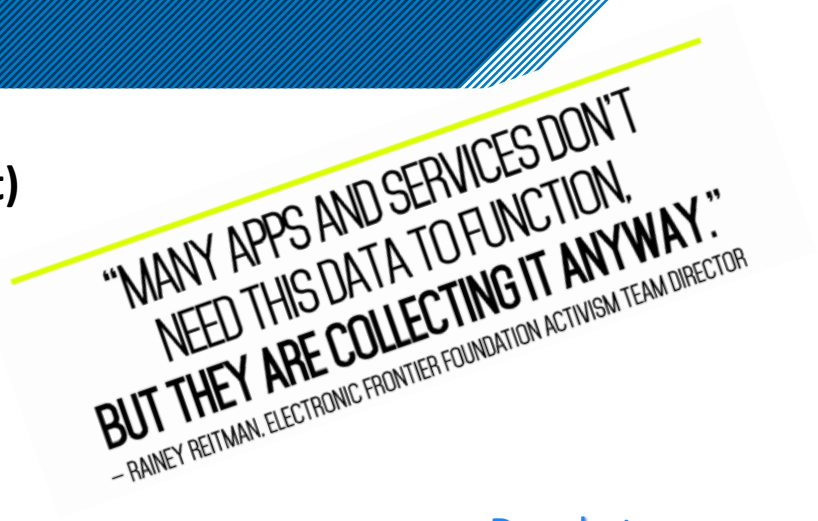


HOW DO WE FIND A BALANCE …..

… between the need for scientific advance in research and care and the fundamental right of each individual to decide in an informed way on the way to live and the amount of privacy

Safeguarding privacy

- The notion of “consent” (informed consent)
- Security
  - Intruder detection
  - Encryption of identifiers
  - Access limitation through roles
  - No need to know the true identity of a subject or center!
    Such a need arises only during data management.
- Certification (NEN7510, ISO27001)
- Transparency
  - Data leak procedures
  - Privacy Impact Assessments



Do whatever you can (technically, financially) even if not strictly required by law

- **The famous trio “necessary”, “proportional”, “subsidiary”**
- **Privacy by Design!**
- **Explanatory memorandum & conscience as guidelines**

HOW TO (MORALLY/LEGALLY) ACCOUNT FOR THE POSSESSION OF PERSONAL DATA

Certification and encryption

- Certification (NEN7510/ISO27001)
  - Health Information Protection

- Encryption
  - TRES, Trusted Real time Encryption Service
  - Via Trusted Third Party

Zorg TTP
Privacy & Vertrouwen

LRQA
Business Assurance

CERTIFICATE OF APPROVAL

This is to certify that the Management System of:

Leiden University Medical Center
(Section Advanced Data Management,
Dep. of Medical Statistics & Bioinformatics)
Einthovenweg 20
2333 ZC Leiden
The Netherlands

has been approved by Lloyd's Register Quality Assurance to the following Standard:

NEN7510:2011
Medische informatica - Informatiebeveiliging in de zorg

The Management System is applicable to:

Design, development, implementation and use of the ProMISe software as well as the design, implementation and maintenance of the infrastructure of medical research data management projects under ProMISe conform the statement of applicability version 2013B.

Approval Certificate No: RQA664586

Original Approval	:	15 March 2010
Current Certificate	:	15 March 2013
Certificate Expiry	:	14 March 2016

Issued by: Lloyd's Register Nederland B.V.

This document is subject to the provision on the reverse
K.P. van der Mandelelaan 41a, 3062 MB Rotterdam, The Netherlands - KvK nr. 24247948
This approval is carried out in accordance with the LRQA assessment and certification procedures and monitored by LRQA.

Trusted Reversible Encryption Service – TRES

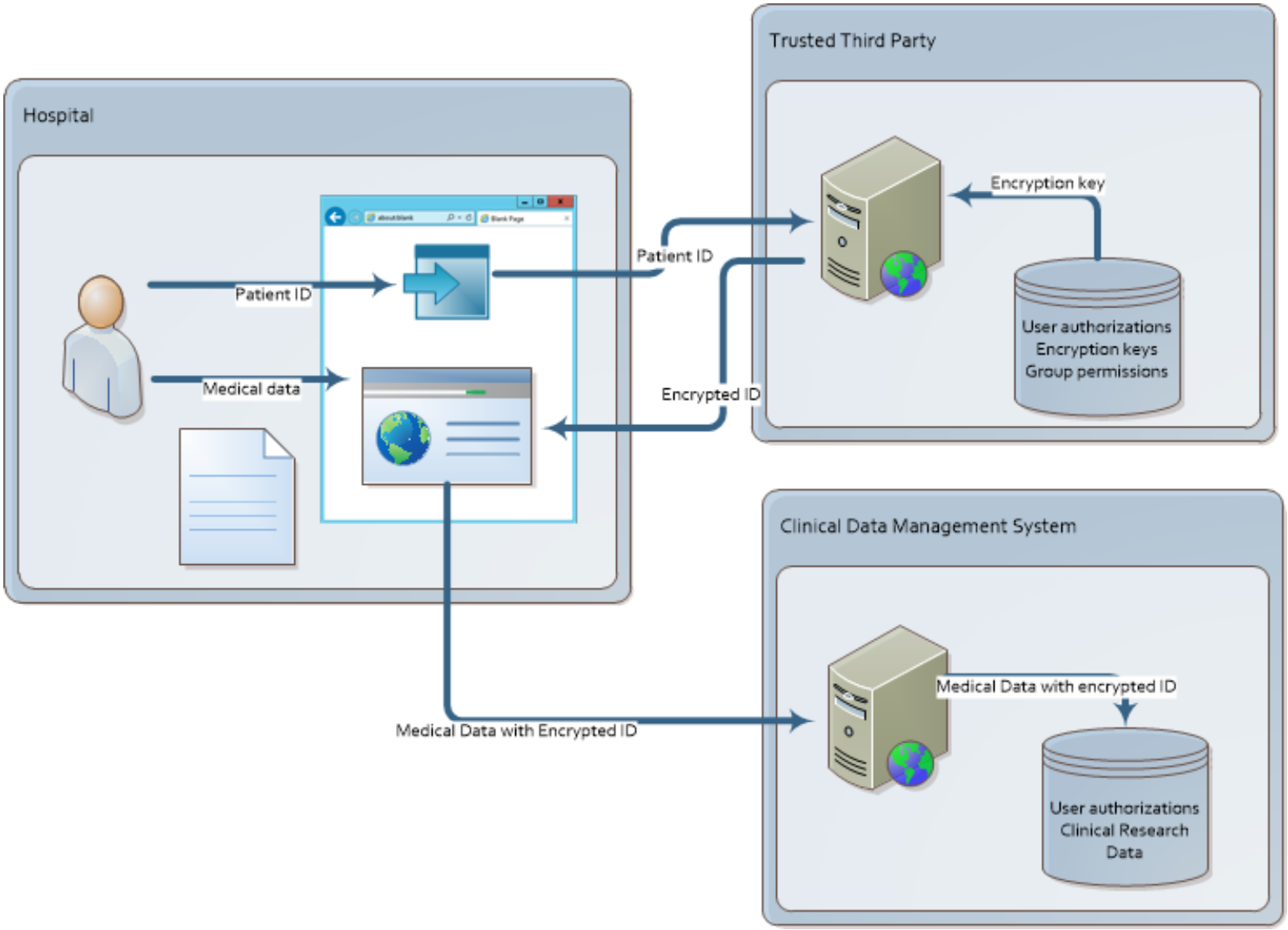
ProMISe

- Transparent real time encryption and decryption
- Based on comprehensive permission system and key management
  - No storage of actual data!
- Supports
  - interactive integration into *any* data management system
  - automated web service/ batch encryption and decryption

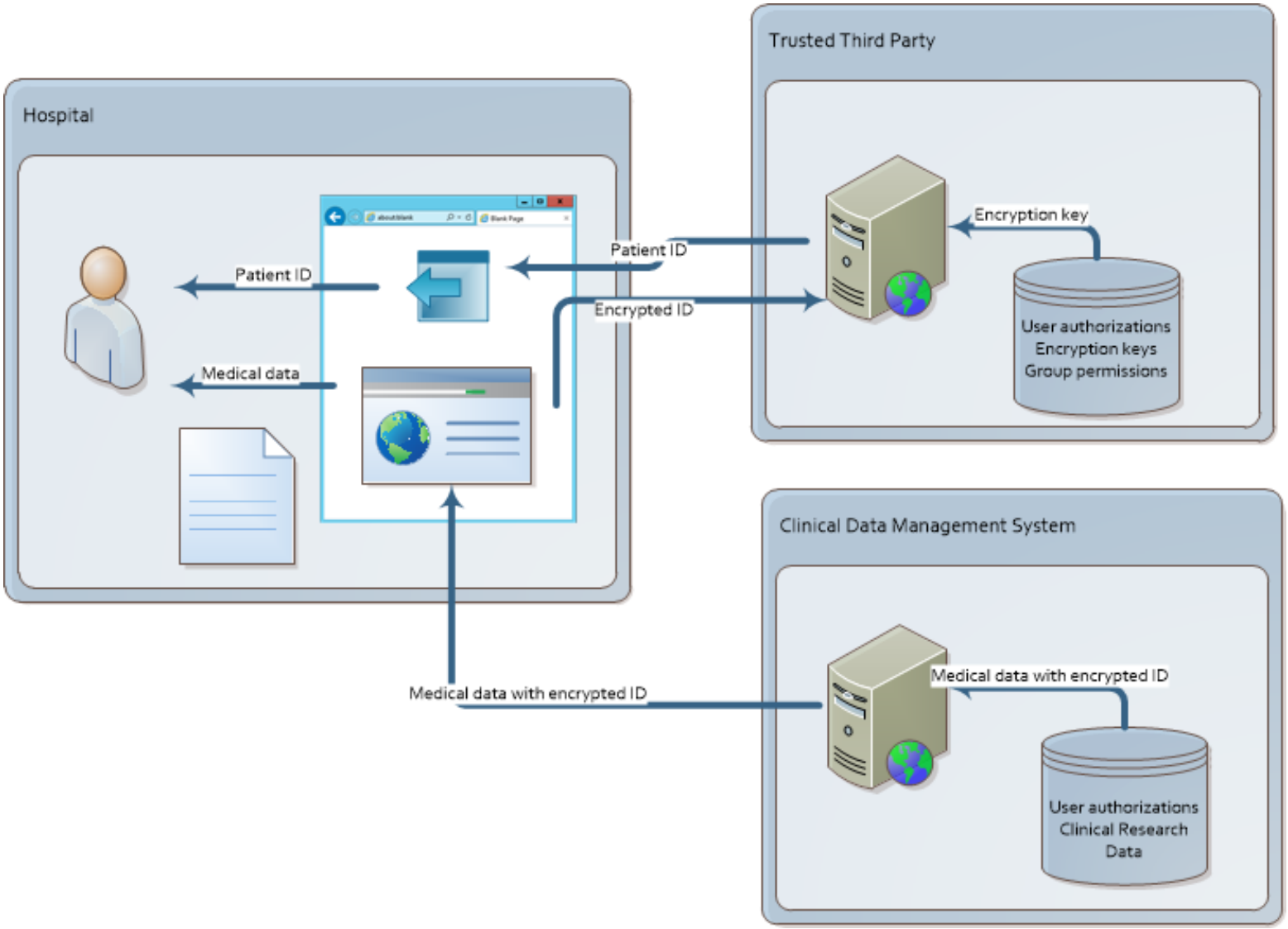
- Invented at the LUMC/ADM and developed in close cooperation with ZorgTTP, a (not-for-profit) Trusted Third Party
- Hosted exclusively by ZorgTPP

Zorg TTP
Privacy & Vertrouwen

TRES integration in (ProMISe) data management: encryption

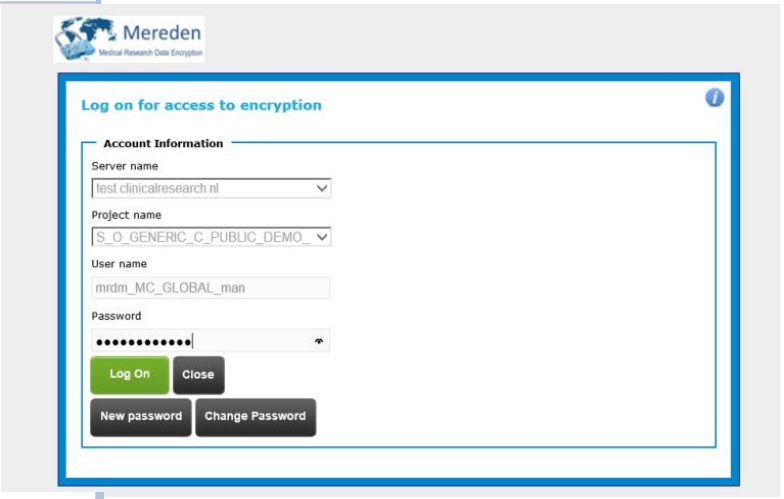


TRES integratation in (ProMISe) data management: decryption



Security

Items for TTP-ENCRYPTION (TOPSECRET)		
Name patient	efcd4d9b-875c-4b2c-ba3 ...	Not connected to TTP
Address patient	efcd4d9b-875c-4b2c-ba3 ...	Not connected to TTP
Social Security Number (BSN)		
Etnicity		



Items for TTP-ENCRYPTION (TOPSECRET)		
Name patient	efcd4d9b-875c-4b2c-ba3 ...	Brand
Address patient	efcd4d9b-875c-4b2c-ba3 ...	Rapenburg 76
Social Security Number (BSN)		
Etnicity		

Items for TTP-ENCRYPTION (TOPSECRET)		
Name patient	efcd4d9b-875c-4b2c-ba3 ...	No decrypt permission.
Address patient	efcd4d9b-875c-4b2c-ba3 ...	No decrypt permission.
Social Security Number (BSN)		
Etnicity		

TRES: generic properties and embedding

- Integrated communication with a Trusted Third Party
- Only the “owner” of a data element can see its original value
- Rights may be extended to others in the same “unit”
- **Searchable encrypted** values allow addition of follow-up data from different locations (in time and space) *without* decryption
- Fully compatible with current legislation on privacy
- On-behalf encryption possible to allow encryption within clusters of hospitals
- Pseudonymized data can be transferred to other domains/organizations
- Completely generic and can just as easily be used in other database systems
- Apart from the “owner”, **nobody (including IT personnel) can infer the original values**

- **Trust by design!**

Possible applications beyond medical research

- Care monitoring at home
- Educational institutions
- Energy sector clients
- Supermarket clients
- Banking clients

Messages

So, in whatever way we collect and share data, in whatever framework and for whatever purpose in whatever IT system, we must remain flexible enough to cope with an ever changing provenance of the data, remain constantly aware of privacy protection requirements and be prepared to apply modern encryption techniques as well as defense mechanisms against unauthorized access of our patient's precious data.

Let's treasure the notion of "privacy in context" ; not privacy as an absolute measure but as something worthwhile to fight for but sometimes not perfectly guaranteed and sometimes not entirely reached. Privacy is a vulnarable entity but so is health. Absolute protection is an illusion and trying to reach it in an absolute sense or with rigid, too specific and a priori established rules contradicts the very notion of risk-benefit assessment by individuals involved.

This was a bit about the way I work and think as a biostatistician who has a Chair in Good Research Data Management…..

But what is your opinion?

"We have to be forthright with the public. We have to have their confidence. We have to convince them we're working for the common good. *Then* we can invade their privacy."