How to Secure Rights of Individuals

ICT 2002,
Regensburg, September 24, 2002

Klaus Pommerening
Institut für Medizinische Biometrie,
Epidemiologie und Informatik
Johannes-Gutenberg-Universität Mainz

Contents

1. Data Protection in Medical Research Networks (TMF)

- Disease specific networks –
  - multicenter clinical trials and epidemiologic research,
  - central data management.
- Important general tasks:
  - Build a network the patient and the physician can trust.
    - Trust in the network should encourage the patient to give his consent for data storage and processing.
  - Show ways how to perform efficient medical research and respect patients' rights.

Basic requirements

- Protect the privacy of an individual's data –
  - Professional discretion
    - maybe the earliest data protection rule in the world,
    - universal validity and acceptance in our culture,
    - protected by criminal and civil law.
  - Data Protection laws.
- Find solutions for security problems by an interdisciplinary and integrative effort:
  - legal, social, methodical, organisational, and technical approaches.

The Rights of individuals

- Medical treatment following the best known practice.
  - The patient's primary interest is not in contributing her data for research projects.
- Privacy of medical and other sensitive personal data.
  - Informed consent for storing and processing data.
  - Revocation of consent without disadvantage.
  - Transparent processing of data.
  - Use of data following the Need-to-Know principle.
- Legal means for prosecuting offences.

Advice by a data protection officer for medical research networks

[B. Sokol (NRW) in her annual report 1999/2000]

- Inform the patient comprehensively about the processing of his data, and get his written consent.
- Make written contracts between participating physicians and the network.
- Work with anonymous or at least pseudonymous data.
- Involve a trusted third party (»Datentreuhänder«) that is protected by law (e. g. notary).
- Don’t use unique patient identifiers across distinct networks.

2. Projects of the TMF working group on Data Protection

- Study data protection scenarios for diverse kinds of research networks ...
  - e. g. for research data base with a clinical focus or in a treatment context, or for epidemiology of infection diseases.
- ... and for communication between distinct networks.
- Pseudonymisation –
  - organisational models, contracts, technical components, consensus with data protection officers.
- Security infrastructure (PKI).
- Framework of policies for medical research networks
  - adopt SEISMED guidelindes and »Grundschutz-Handbuch« (Basic IT security manual) of BSI
- ...

Example: Security infrastructure

[A working security infrastructure is an essential base for a trusted network.]

- Build a PKI and interface it with system components.
- Simple and cheap approach:
  - Use PGP for personal communication,
  - Use SSL and passwords for client-server communication.
- State-of-the-art approach: Smartcard based PKI.
- Obstacles:
  - Proprietary solutions violate standards.
  - No easy integration into existing system components.
- First implementations this year.

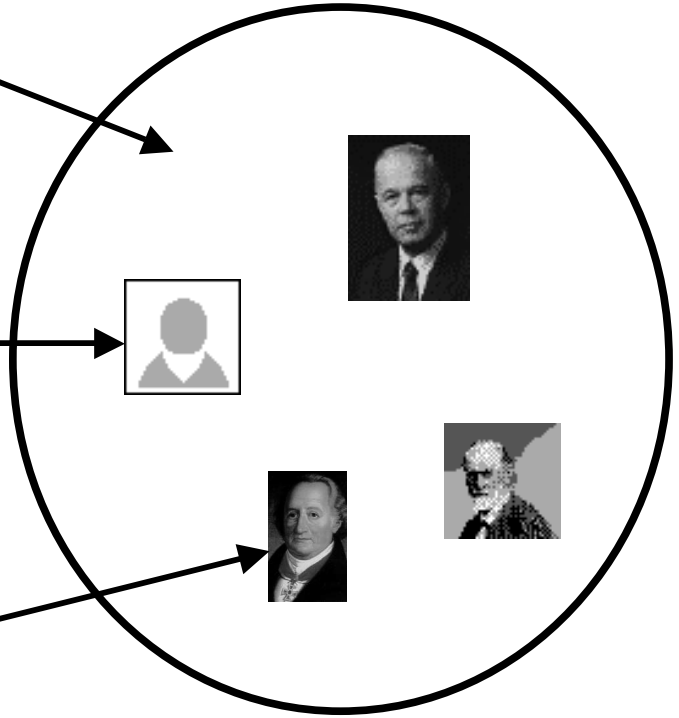
3. Pseudonyms

- The golden mean between anonymous data and identity (or identity revealing) data.
- Almost as good as anonymity, depending on context –
  - one-way pseudonyms can’t be reversed,
  - reversible pseudonyms can.
- Almost no restraints for data processing:
  - record linkage from different sources possible,
  - feed back to the patient possible, depending on organisational and technical framework.
- Feasability!? Expenses!?

Anonymous record
points to a set of individuals

Pseudonymous record
points to a single individual whose identity is concealed

Identifyable record
points to a specific individual.



Identity, pseudonymity, anonymity

- Anonymous records:
  - There is no way back to the identity (for feedback, for follow-up information)
  - and no way for linking records.
- Pseudonymous records
  - are distinguishable,
  - can be linked,
  - but can’t be associated with a specific person,
  - can be associated under strong restrictions only, if provided for by legal, organisational, and technical means.

Pseudonyms for medical research I

Pseudonymisation is an important instrument for medical research.

Principal rules (as a first approxmation):

- In a clinical (treatment) context the identity of the patient is needed and allowed.
- In a research context anonymous records are allowed, and should be used whereever possible;
- otherwise use pseudonymous records.

Pseudonyms for medical research II

- One-way pseudonyms are »de-facto anonymous«.
- Reversible pseudonyms are not: written consent of the patient is necessary.
- Quality assessment may be viewed as part of the treatment context, however the use of pseudonyms should be considered.
- Always use the method that diminishes privacy as little as possible.

Early examples of pseudonymisation

- Untraceable electronic money (Chaum ca 1980) [implementation withdrawn].
- Electronic prescription (Struif ca 1990), pseudonymous settling of bills in health care (GMDS-AG DS) [never realized].
- Cancer registry (Michaelis/P. 1993).
- Consensus paper »Epidemiologie und Datenschutz« (DAE/ AK Wissenschaft der DSB).
- QuaSiNiere project.

Several recent german laws require pseudonymisation in appropriate contexts.

The TMF pseudonymisation project

- Preparatory work in single networks –
  - KN POH: Pseudonyms for communication in multicenter trials and in cancer registry.
  - KN CED/Leukämie: Pseudonymisation during export of data for research projects.
  - KN Parkinson/Rheuma: Pseudonymous central research data base.
- Project DS 3.1 of TMF –
  - reusable organisatorial models with a generic kernel and possible variations,
  - technical components (services),
  - consensus with data protection officers.

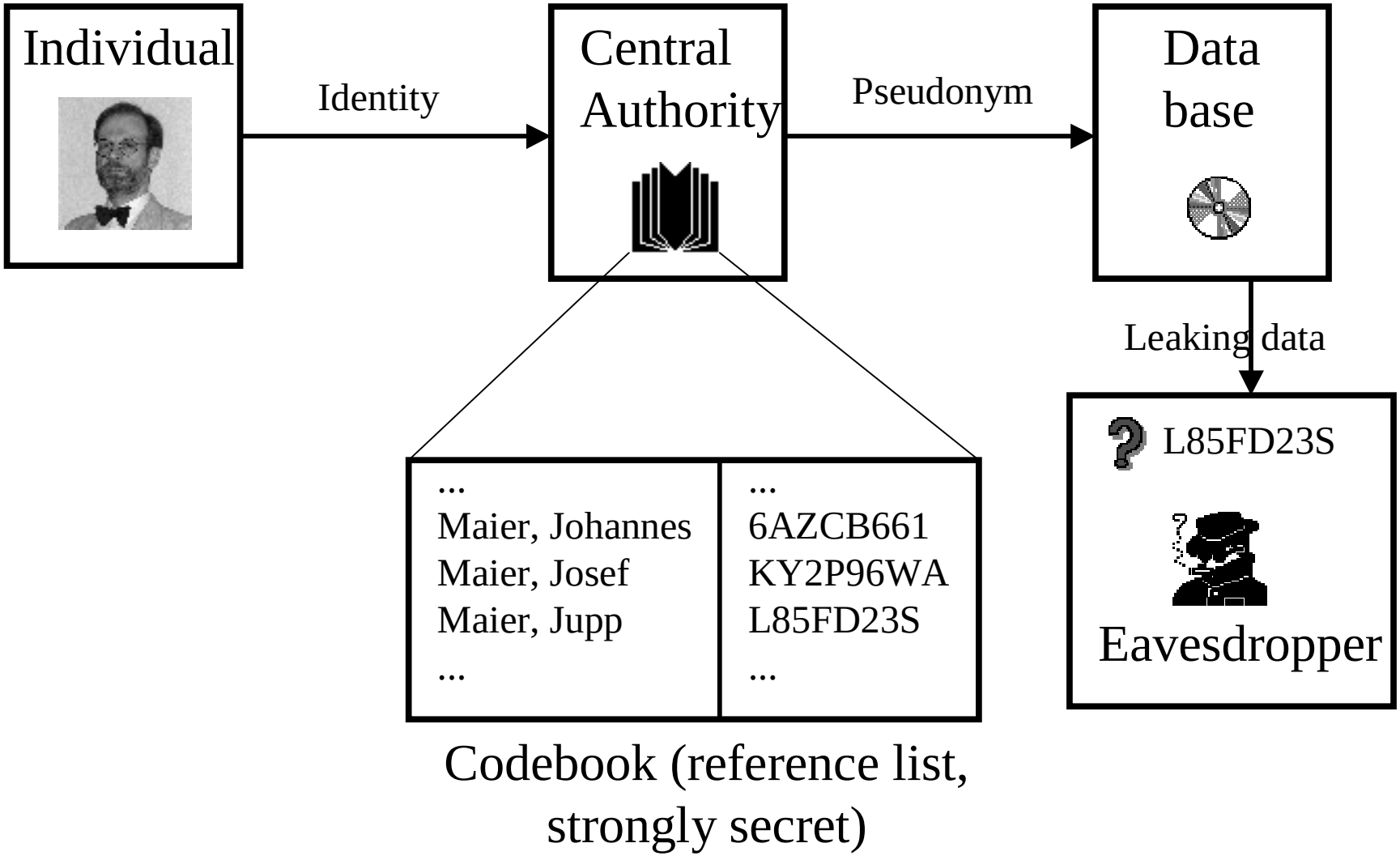
Example: Database in a clinical environment (Reng)

- Central database inside a participating hospital
  - treatment data for patients of participants of the (disease specific) network
  - with strong access control
  - contractual framework.
- Research with pseudonymous data
  - either one-way pseudonyms (allow record linkage),
  - or reversible pseudonyms (allow feed back of important results, e. g. genetic dispositions)
  - depending on needs of project.

4. Concepts for pseudonymisation

- Organisatorial criteria:
  - Who »owns« the pseudonym? Who creates it? Who can reveal it?
  - When introduce the pseudonym? [e. g. quality assurance of data before pseudonymisation!]
- Technical criteria:
  - How to construct the pseudonym?
    [by random or by a deterministic procedure?]
  - How to guarantee the unique association?
  - How to enable or prevent re-identification?

A (too) simple model: The codebook



The codebook model

- The central authority generates *and stores* the pseudonyms.
- ... can re-identify without co-operation with the individual,
- ... is an outstanding target,
- ... has to be absolutely trusted by all parties.
  - TTP = Trusted Third Party.
- ... provides reversible pseudonyms only.

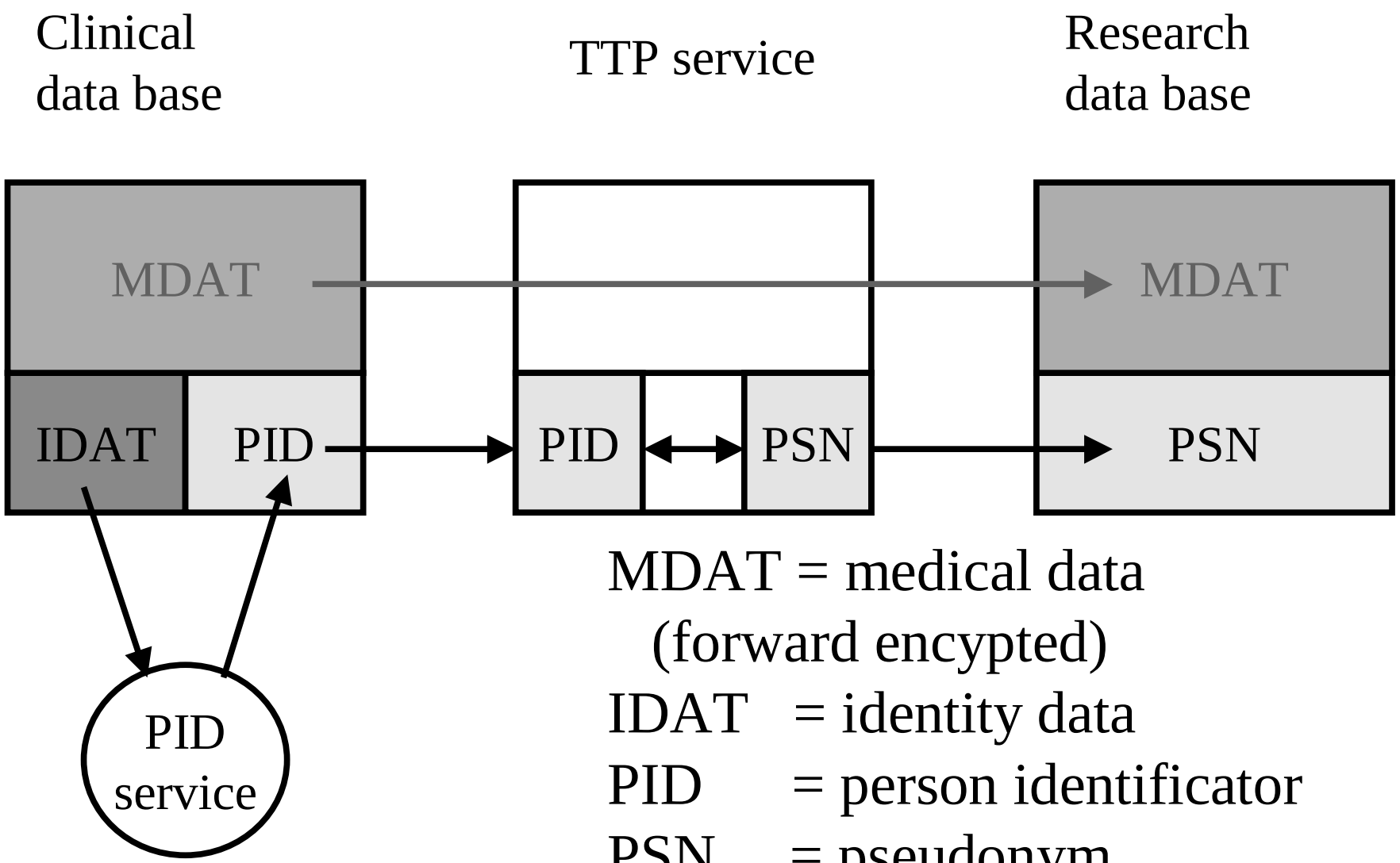
Goals for better solution:

- Minimize trust assumptions.
- Split informational power.

The TMF TTP model I

- Approach: Deterministic procedure for generating pseudonyms –
  - obsoletes reference list.
  - Method 1: Key-dependent hash value from identity data.
  - Method 2 (even better, guaranteed uniqueness): encrypted person identificator (PID).
- The TTP stores nothing but its secret key.
  - Several threats fall away.

The TMF TTP model II



MDAT = medical data
(forward encypted)
IDAT = identity data
PID = person identificator
PSN = pseudonym

The TMF TTP model III

- TTP service:
  - Encrypt PID into PSN (and vice versa, if applicable),
  - MDAT and IDAT unknown.
- Clinical data base:
  - MDAT, IDAT, PID known,
  - PSN unknown.
- Research data base:
  - MDAT and PSN known,
  - IDAT and PID unknown.
- PID service (patient registry):
  - IDAT and PID known, in permanent data base,
  - store IDAT one-way encrypted, if appropriate.

Technical components

- PID service:
  - unique identification service as essential part (solve matching problem),
  - cryptographic procedure with secret key (unique for each distinct network),
  - patient registry.
- Pseudonym service:
  - simple cryptographic procedure with secret key).
- Procedures for communication in intranet
  - using the security infrastructure (PKI),
  - web based via SSL.

The Matching Problem

Without a reliable procedure for record matching pseudonymisation becomes useless.

- Logical Matching
  - recognise (e. g.) name change, parts of names.
- Minimize homonym and synonym errors.
- Error tolerance:
  - Use additional data and phonetic codes,
  - use stochastic procedure,
  - warn user, if appropriate.

The matching problem can be solved in a satisfying way.

The TMF pseudonymisation service

provides models and procedures

- *at first for research networks*,
- *but also for health care and quality assurance*,

that comply with the requirements of professional discretion and data protection in a distinguished way,

and nevertheless allow useful and pertinent processing of medical information.

[... and allows record linkage between different networks, if legal conditions are settled.]

5. The Future: Architecture of a secure medical network

- Use existing basic infrastructure (PKI, best available security tools).
- Build a high level architecture consisting of standardized, easy to use TTP services together with the corresponding contractual framework und policies.
- Define and implement the needed legal and organisational structures in a european setting.
- Create sustainable, lasting structures with the help of industrial partners.
- Support close co-operation between health care providers and research groups in an efficient and secure way.

Example: Web based services

based on XML, CORBA, SOAP (for pilots see HARP project, Blobel):

- Policy services –
  - policy definition, interpretation, enforcement, bridging and mapping.
- Access control services –
  - directory, authentication, management of roles and rights, access decision, proxies and gateways through firewalls and beyond policy borders.
- Accountability/ liability services –
  - notary functions, timestamp services, component certification, auditing, monitoring, detection of illegal behaviour.
- ... and other TTP services.