Secondary Use of the EHR via Pseudonymisation

Klaus Pommerening, Mainz

Michael Reng, Regensburg

TMF = Telematikplattform für die medizinischen Forschungsnetze

[Telematics Platform for the German Health Research Networks]

Contents

1. **EHR and Pseudonyms**
2. 5 Scenarios for Secondary Use
3. Results
4. Discussion

Uses of the EHR

- Primary use: Treatment context.
- Secondary uses:
  - Disease specific clinical or epidemiological research projects,
  - Health care research, assessment of treatment quality, health economy.

Typical aspects of secondary uses:

- Outside of treatment context and professional discretion (of the treating physician),
- The identity of the patient doesn't matter.

For secondary use of the EHR:

- Protect the identities of the patients.
- Anonymisation whereever possible.
  Drawbacks of anonymisation:
  - No association between data from distinct sources
  - ... or from distinct points of time.
  - No way back to the patient for feedback
  - ... or for recruiting suitable patients for a new research project.

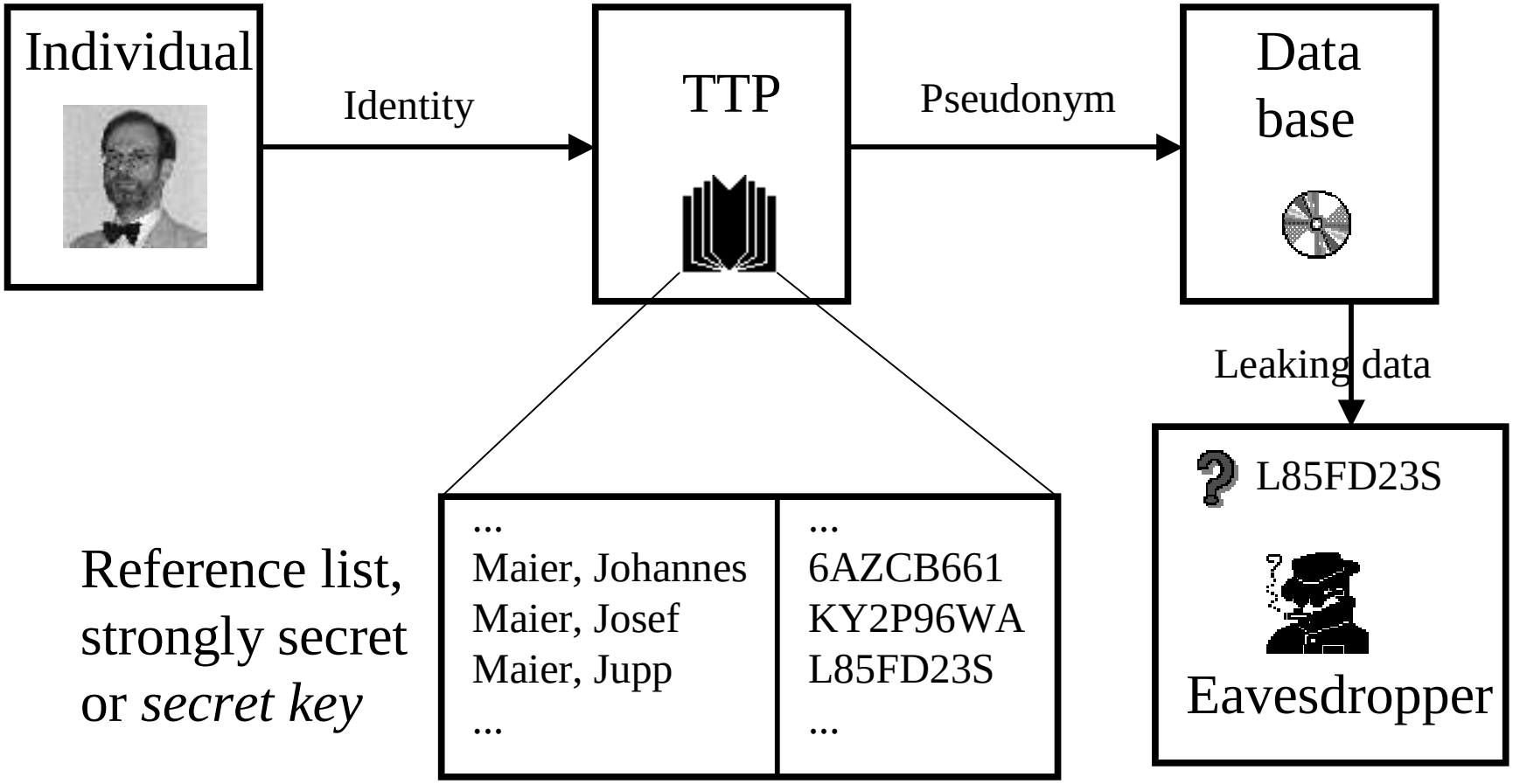
Pseudonyms

- 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 allow re-identification of the individual.
    - Written informed consent necessary for reversibility!

Basic Types of Pseudonyms

- Untraceable pseudonyms (Chaum ca 1980)
  - Based on blind digital signature,
  - Under control of owner,
  - Not suited for secondary uses of the EHR.
- TTP-generated pseudonyms
  - Trusted Third Party = »Vertrauensstelle« or »Datentreuhänder« (e. g. a notary).
  - Example: Cancer registry (Michaelis/Pomm. 1993).

TTP-generated Pseudonyms
(Basic Model)



Contents

1. EHR and Pseudonyms
2. **5 Scenarios for Secondary Use**
3. Results
4. Discussion

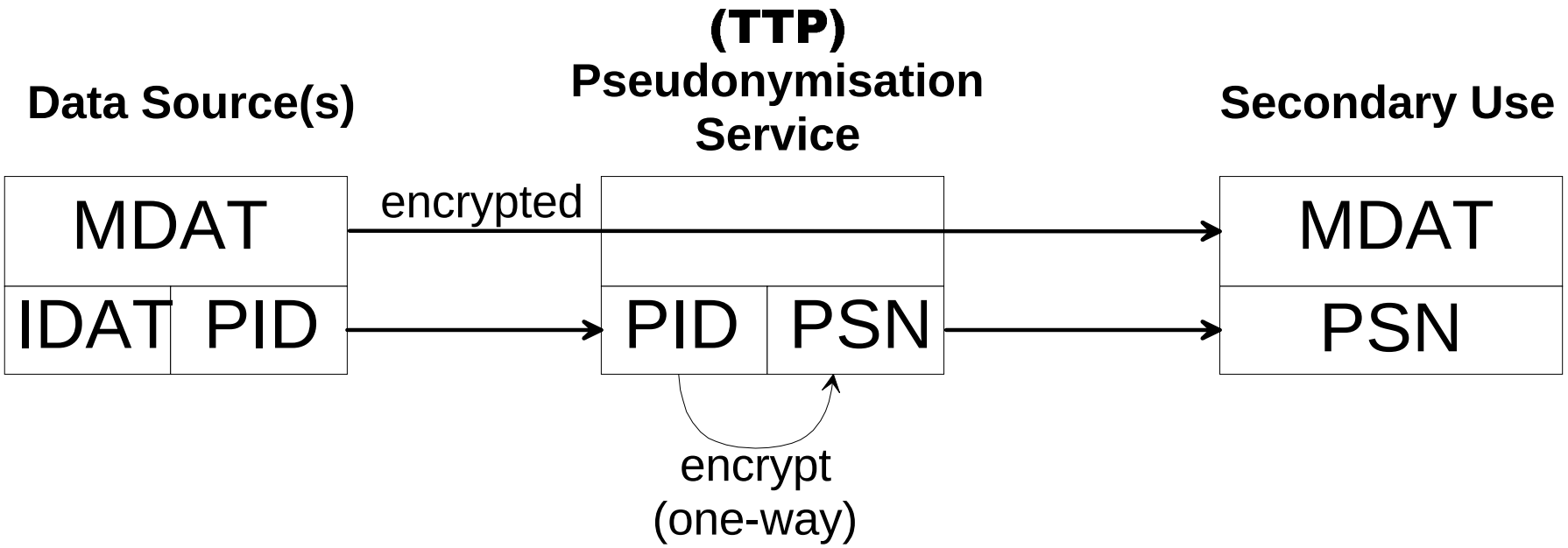
(1) Single Data Source, One-Time Secondary Use

- Typical application case for anonymisation.
- Example: A simple statistical evaluation of EHR data.

(2) Overlapping Data Sources, One-Time Secondary Use

- Data from diverse sources must be linked together.
- Examples:
  - Multicentric study,
  - Follow-up data.
- Typical application case for one-way pseudonyms.

Pseudonymisation for One-Time Secondary Use



MDAT = Medical Data
IDAT = Identity Data

PID = Unique Patient Identifier
PSN = Pseudonym

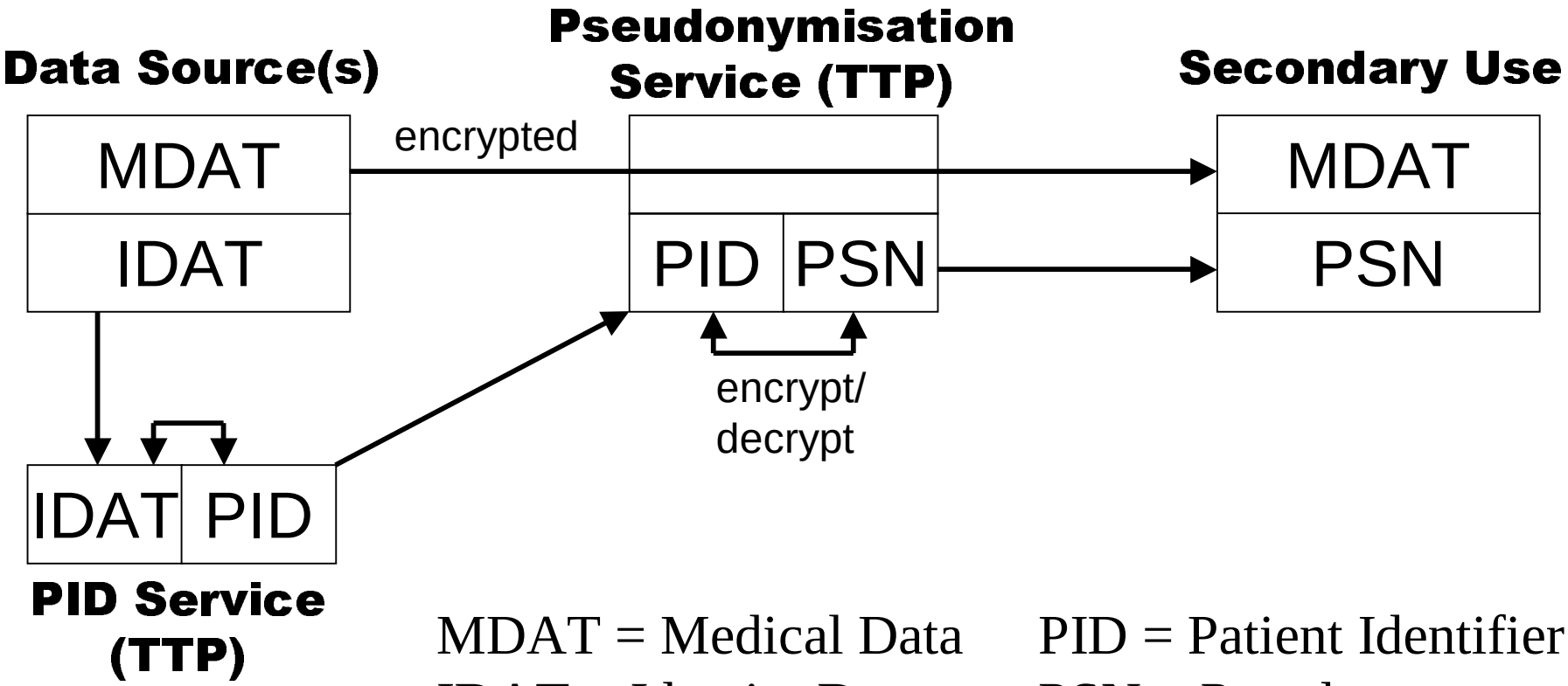
Properties of Scenario (2)

- Medical data (MDAT) are encrypted with public key of secondary user –
  - The TTP cannot read the MDAT.
  - Only the secondary user can decrypt them.
- The pseudonym (PSN) is the encrypted PID
  - With a secret key, known only to the TTP,
  - By a one-way procedure.
- The TTP doesn’t store anything (except the key).

(3) One-Time Secondary Use with Re-Identification

- Use the »Basic TTP« model,
  - But no reference list, only secret TTP key.
  - PSN service performs reversible encryption procedure.
- Use a non-public (project specific) PID
  - Generated by a separate TTP service.
  - PID service stores association between IDAT and PID (»Patient List«).
- Re-identification involves PSN service and PID service.

Pseudonymisation with Possible Re-Identification



MDAT = Medical Data
IDAT = Identity Data
PID = Patient Identifier
PSN = Pseudonym

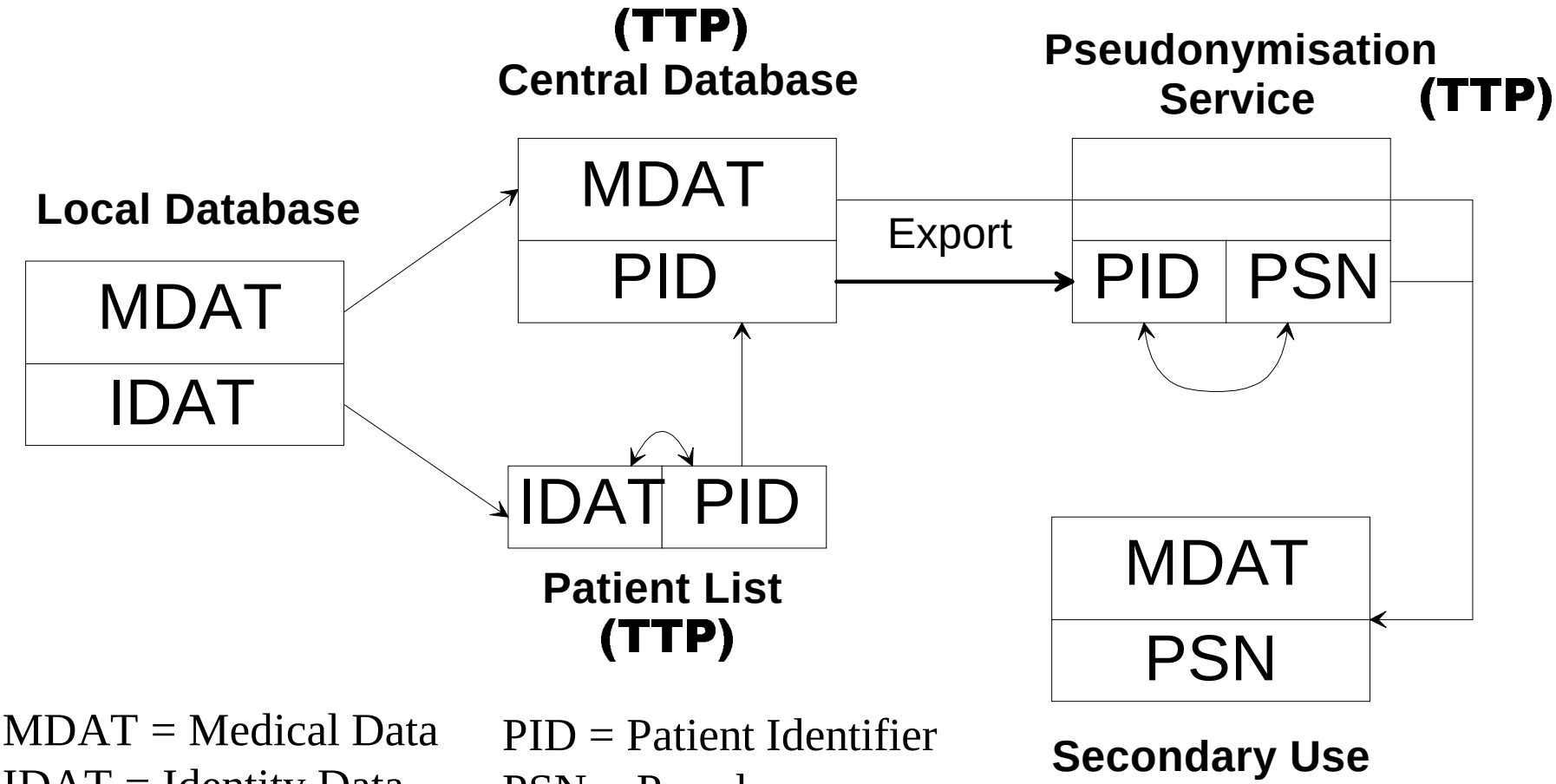
(4) Pseudonymous Research Data Pool

- Same procedure as in (3),
  - But the secondary user builds a (disease specific) registry.
- Long term data accumulation needs special organisational and technical security measures.
- Quality management of data should precede pseudonymisation.
  - Yet another TTP service.
- »Model B« of the generic concept of the TMF.

(5) Central Clinical Data Base, Many Secondary Uses

- Data pool = central »clinical« data base.
  - Access for treating clinician (decentral).
  - No identity data, only PIDs.
  - Access by temporary tokens.
  - Implemented as (yet another) TTP service.
- No online access by secondary users.
  - Secondary users get exported data set (anonymised or pseudonymised).

TTPs for Central Clinical Data Base



MDAT = Medical Data
IDAT = Identity Data

PID = Patient Identifier
PSN = Pseudonym

Properties of Scenario (5)

- Advantages:
  - Better support for long-term observation of patients with chronic diseases.
  - Useful for the data producing clinician.
  - Individual feedback of research results easy.
  - Fits well into EHR architecture.
- Drawback:
  - Sophisticated communication procedures.
  - More TTPs and secret keys involved.

»Model A« of the generic concept of the TMF

Contents

1. EHR and Pseudonyms
2. 5 Scenarios for Secondary Use
3. **Results**
4. Discussion

Results I

- TMF models A and B [(5) and (4)] approved by the German Data Protection Commissioners
  - (Arbeitskreis Wissenschaft der Datenschutzbeauftragten des Bundes und der Länder)
- Scenario (2) in routine use since 2002 in a health care research project of the TMF.
- Scenario (5) implemented in a research network.
  - KN CED (Chronic Inflammatory Bowel Disease).
  - Further implementations in progress.

Results II

- Scenario (4) adapted by several research networks
  - Implementations in progress.
- TMF offers software tools for the TTP services.
- Corresponding policies, sample contracts, forms for patient's consent available from TMF (free for members).

Contents

1. EHR and Pseudonyms
2. 5 Scenarios for Secondary Use
3. Results
4. **Discussion**

Discussion I

- The TMF model architecture (variants A and B) provides ways for building central data pools for medical and health care research, that
  - conform to the German and European data protection rules,
  - respect the patients' rights,
  - and cover a wide range of situations.
- The pseudonymisation scenarios look complex, but once established, work transparently.

Discussion II

- The transfer to other applications in health care is possible and recommended.
  - We could – and should – build TTP services for secondary uses into the EHR architecture (suitably adapted from the TMF services).