

Integrating eHealth and Medical Research: The TMF Data Protection Scheme

Klaus POMMERENING^{a,1}, Ulrich SAX^b, Thomas MÜLLER^c, Ronald SPEER^d,
Thomas GANSLANDT^e, Johannes DREPPER^f, and Sebastian C. SEMLER^f
^a *Johannes-Gutenberg-Universität, Mainz, Germany*
^b *Georg-August-Universität, Göttingen, Germany*
^c *Ludwig-Maximilians-Universität, München, Germany*
^d *Universität Leipzig, Germany*
^e *Friedrich-Alexander-Universität, Erlangen, Germany*
^f *TMF e. V., Berlin, Germany*

Abstract. Building medical research networks is an important part of eHealth. Medical research advances diagnostic and therapeutic knowledge and standards and thereby benefits the patients. On the other hand patient data and samples are among the most sensitive personal informations and must be carefully protected according to rules of ethics and professional discretion as well as national and international data protection laws. The TMF therefore developed a “generic” scheme with variants for the processing of information in networks and biobanks that respects the patients’ rights on privacy. The main methods are: separation of informational powers and responsibilities, pseudonymisation, templates for the informed consent, policies, and contracts. A major revision of the scheme, taking into account the experiences from completed or ongoing implementations, is in progress.

Keywords: Medical Research, Data Protection, Databases, Pseudonyms

1. Introduction

1.1. Healthcare and Research

Interconnecting medical research with healthcare is a central challenge in eHealth, in particular to the end of adjusting medical treatment to the latest research findings. For this we need an efficient transfer of knowledge from the sphere of research into that of healthcare; in the other direction medical research needs reliable data and samples from patients. Thus the progress of medicine crucially depends on the smooth transfer of ideas, questions, and data between healthcare and research; an efficient common infrastructure is a precondition for this. Moreover medical research can be successful and internationally competitive only with interdisciplinary, interregional or international collaboration.

¹ Corresponding Author: Klaus Pommerening, Institut für Med. Biometrie, Epidemiologie und Informatik, Johannes-Gutenberg-Universität, 55101 Mainz, Germany. <http://www.staff.uni-mainz.de/pommeren/>

1.2. Examples of Research Oriented Medical Networks in Germany

As a typical example look at a scenario from Pediatric Oncology and Hematology, where collaboration between healthcare and research is organised in the KPOH [1] (Competence Network for Pediatric Oncology and Hematology): A pediatrician suspects a malign disease in a girl. He refers the child to a special hospital. There a specialist ensures the diagnosis with the help of central reference institutions (pathology, radiology), admits the patient to the corresponding multicentric therapy study, and treats her according to the study protocol. The principal investigator of the study acts as supervisor of the therapy. The treating hospital transfers the study data to the study center, and a basis data set to the national Childhood Cancer Registry, that serves as epidemiologic registry. Moreover tissue or blood of the patient is given to a study specific biobank or a central tumor registry; in many cases also image data are stored in an image database for reference purposes. The clinical studies aim at optimizing the therapy plans, and so immediately improve healthcare. The data in the cancer registry are the basis for studies on late effects, secondary malignoma, and quality of life, that influence healthcare in the future.

The situation in the KN AHF [2] (Competence Network for Congenital Heart Disease) is roughly similar; however this network maintains a central study database instead of separate data bases for the single studies, and it has a research database, where study data are kept for a long time, in addition to a registry, that maintains a lifelong basic patient record but no detailed research data.

1.3. Architecture of a Research Network

A typical medical research network contains some or all of the following components, sometimes even in multiple instances:

- Patient database for immediate healthcare,
- Clinical database for inter-institutional accumulation and evaluation of treatment data,
- Study database for the central data management of clinical studies,
- Research database for long-term accumulation of research data,
- Registry,
- Medical image database,
- Biobank,
- Electronic archive for long-term conservation of data that are no longer needed for treatment or research but must be kept for documentation purposes.

There are obvious data flows between all these databases, and therefore a research network has a quite sophisticated architecture. Modelling the processes in the network is a demanding task and starts with the specification of use cases, each giving an external view at a database and its connections and boundary conditions. Moreover a data protection concept also needs an internal view at the details of the underlying business processes.

1.4. The Telematics Platform TMF

In the TMF [3] (Telematics Platform for Medical Research Networks), medical research associations work together in identifying common issues and solving technical, legal, and organisational problems about interconnection of research and healthcare, standards and terminology, legal and ethical frameworks, quality management, technology assessment, public relations.

The TMF members include the Competence Networks in Medicine, the Coordinating Centres for Clinical Trials (KKS), networks dealing with rare diseases, infection epidemiological networks, the National Genome Research Network (NGFN), and various other networked medical research organizations.

2. Data Protection in Healthcare and Research

2.1. Some Principles

Patient data and biomaterials are among the most sensitive personal informations and must be carefully protected according to rules of ethics and professional discretion as well as national and international data protection laws.

In healthcare we primarily have a treatment context. Here the patient is—and should be—personally known by name. Data protection mainly follows the rules of professional discretion but also—subsidiary—the data protection laws [4].

However the treatment and research contexts must be separated carefully. Typical aspects of secondary uses of patient data [5], such as medical research, are that

- the data leave the treatment context,
- the identity of the patient doesn't matter, there is no direct contact.

In such a context use of anonymous data is allowed; therefore anonymisation should be performed whenever possible. But this doesn't always work: In many cases of medical research the correct association between a single patient's data from distinct sources or distinct points of time is crucial. In some scenarios even a way back to the person is required: It could be in the interest of the patient to learn about results of a research project, for example a genetic disposition; or a researcher might want to use a data pool to recruit suitable patients for a new clinical or epidemiological study.

2.2. Methods and Tools for Data Protection

Using data from healthcare for medical research, and respecting patients' rights and the principles of ethics and data protection, requires the careful use of suitable methods and tools, such as:

- Separation of informational power, duties and responsibilities, for example separate administration of distinct databases.
- Pseudonymisation at least at the border between healthcare and research, in many cases even multiple pseudonymisation.
- Checklists and templates for the informed consent [6].
- Policies and contracts that define the duties of the network staff and of external partners.

2.3. The Generic Data Protection Scheme

The Generic Data Protection Scheme [7] of the TMF showed ways for successful research respecting the data protection requirements by using the methods and tools listed above. The first version introduced two different network types, A and B, which were adapted to two different (typical but specific) research networks. However the concept almost nowhere addressed the integration of healthcare and research.

The scheme was developed in close collaboration with the German Data Protection Commissioners. Some research networks already adapted it; several other networks are in the processes of planning or implementation.

2.4. Biobanks for Medical Research

The handling of samples has some analogies with the handling of data, but there are also a few differences, in particular is the full genetic information contained in each piece of material. Therefore samples cannot be assumed anonymous except for a quite narrow future. This requires careful use restrictions, physical separation of data and samples, and separate databases for analysis data and other medical data. The TMF introduced a third network type ('type BMB'), in fact developed a modular scheme that comprises many different variants of biobanks [8].

3. Evaluation of the Generic Scheme

3.1. Workshop

Several German medical research networks (more than 20 up to 2006) adapted their data protection plans from the generic TMF scheme. To evaluate their experiences, the TMF conducted a workshop. We collected statements on specifics of the approach, judgements of the data protection commissioners, problems with implementation and operation, as well as wishes and proposals for improvements.

The main lessons learnt were: Deriving a specific plan from one of the generic types requires much work, and the implementation is tedious. Several requirements had no counterpart in the generic models, for example the needs of a network with many clinical multicenter studies, or the tight integration with healthcare.

3.2. Requirements for a Revised Version

Many networks need a mix of both generic types A and B: Type A for accompanying chronically ill patients over a long period, or for the central management of clinical studies, or for quality assurance of data; type B for building a long-term data pool, or an epidemiologic registry; and moreover type BMB for their biobanks. So we identified a list of requirements for a revision of the generic scheme:

- Comprehensive specification of use cases and business processes.
- Replacing the distinct types A and B by a modular approach, distinguishing databases in the treatment context from research databases with potential

feedback to the treatment process, and from databases in a pure research context.

- Scalability of measures according to criteria of adequacy depending on the size of the network after assessing the risk of re-identification.
- Integration of healthcare and research structures.
- Better specification of the processes of quality assurance.
- Integration of clinical multicenter studies, registries, and image data bases.
- Proposals for central services.

Moreover we identified a number of open legal questions to be answered after a comprehensive review of the legal situation. The TMF project group has almost finished the revision, and will then again seek the approval of the data protection commissioners. After that the text will be published as a book.

4. Conclusion

The Generic Data Protection Scheme of the TMF, in particular in its revised version, is a flexible and scalable basis for cooperation between research and healthcare, supports long-term data accumulation for research as well as for healthcare, and enables all kinds of secondary uses of patient data and samples.

A medical network or research project may use the generic scheme as a template, choose the appropriate components, and simplify the generic architecture according to project specific needs following the criteria of adequateness. The TMF will assist the network in the process of getting the approval by the competent Data Protection Commissioner.

The generic TMF architecture is not a static structure. There are practical experiences and feedback from implementations, but also changing requirements and developing ethical frameworks in health care research and medical networks, for example with respect to genetic research. Therefore the TMF must continually keep its generic scheme up to date to meet new challenges.

Acknowledgement

This work was supported by the Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF) as a project of the TMF (Telematikplattform für Medizinische Forschungsnetze e. V.). The first author is chair of the TMF Working Group on Data Protection and representative of the KPOH (Kompetenznetz für die Pädiatrische Onkologie und Hämatologie) in this group. The other authors represent the Competence Networks “Angeborene Herzfehler” (Fördernummer 01 GI 0201), “Leukämien”, “Hepatitis”, “Maligne Lymphome”, the KKS Leipzig, and the Network “Epidermolysis Bullosa”.

References

- [1] Kompetenznetz Pädiatrische Onkologie und Hämatologie. Online: <http://www.kinderkrebsinfo.de/> (last accessed March 15, 2007).
- [2] Kompetenznetz Angeborene Herzfehler. Online: <http://www.kompetenznetz-ahf.de/> (last accessed March 15, 2007).
- [3] Telematik-Plattform für Medizinische Forschungsnetze e. V. Online: <http://www.tmf-ev.de/> (last accessed March 15, 2007).
- [4] Pommerening K, Blobel B. Datenschutz und Datensicherheit in Informationssystemen des Gesundheitswesens. *Führen & Wirtschaften im Krankenhaus* 2/1997, 133-138.
- [5] Pommerening K, Reng CM. Secondary use of the Electronic Health Record via pseudonymisation. In Bos L et al. (eds.). *Medical Care Compunetics I*. Amsterdam: IOS Press; 2004; pp. 441-446.
- [6] Harnischmacher U et al. Checkliste und Leitfaden zur Patienteneinwilligung. München: MWV; 2006.
- [7] Reng CM et al. Generische Lösungen zum Datenschutz für die Forschungsnetze in der Medizin. München: MWV; 2006.
- [8] Pommerening K et al. Ein generisches Datenschutzkonzept für Biomaterialbanken. München: MWV; 2007.