DZHK Central Data Management Process Description and Data Protection Concept

DEUTSCHES ZENTRUM FÜR HERZ-KREISLAUF-FORSCHUNG E.V.
(GERMAN CENTRE FOR CARDIOVASCULAR RESEARCH)

Version 1.2, 24 March 2014

Published by:

University Medicine Greifswald
Institute for Community Medicine

represented by the Managing Director
Prof. Dr. med. Wolfgang Hoffmann, MPH

Ellernholzstr. 1-2
D-17487 Greifswald

Tel: +49 (0) 3834 86 7750
Fax: +49 (0) 3834 86 7752
email: wolfgang.hoffmann@uni-greifswald.de

University Medical Centre Göttingen
Department of Medical Informatics

represented by the Director
Prof. Dr. med. Otto Rienhoff

Robert-Koch-Straße 40
D-37075 Göttingen

Tel: +49 (0) 551 39 3431
Fax: +49 (0) 551 39 22493
email: otto.rienhoff@med.uni-goettingen.de

Content

A Project-specific section

A1 CDM project overview

1 Project structures and partners

The following section introduces the framework of the project in which the concept of the Central Data Management (CDM) unit is to be implemented within the German Centre for Cardiovascular Research (DZHK). The CDM unit consists of the independent Trusted Third Party (TTP) and the Data Handling (DH) unit. It will also explain the project-specific requirements and the resulting measures.

1.1 DZHK

The DZHK is one of the six German Centres for Health Research (DZGs). Their foundation was initiated by the German Federal Ministry of Education and Research (BMBF) and took place between 2009 and 2012. The health centres are organised as "eingetragene Vereine" ("e. V."), which are registered associations under German law. They are registered as part of the Helmholtz- Gemeinschaft Deutscher Forschungszentren e.V. (Helmholtz Association of German Research Centres), making them research networks within Germany's largest non-university scientific organisation. The intention is that the health centres work in close cooperation with one another in order to achieve their goals faster.

The aim of the DZHK is to improve the prevention, diagnosis and treatment of heart and cardiovascular diseases, and to ensure by way of wide-ranging cooperation that research results from this area of expertise find their way more quickly into clinical practice.

To achieve this, 27 partner institutions are working at the following seven locations / in the following location groups within the DZHK: Berlin, Göttingen, Greifswald, Hamburg-Kiel-Lübeck, Heidelberg-Mannheim, Munich and Rhine-Main (Frankfurt a.M., Bad Nauheim, Mainz).

These partner institutions are universities, university hospitals and non-university research institutions (several Max Planck Institutes, the Max Delbrück Centre and one Leibniz Centre). The plan is to cooperate with strong partners from the German healthcare sector.

The work of all the health centres is to be overseen and assessed by high-level advisory bodies staffed by international consultants. The aim is to evaluate not only scientific excellence, but also the strategic direction and the established structures and processes.

As with the other DZGs, the DZHK is 90 % financed by the German federal government. The remaining 10 % of the financing come from the federal states where the respective health centre maintains sites (in the case of the DZHK this includes nine federal states). [1]

1.2 Institute for Community Medicine

The focus of the Epidemiology of Health Care and Community Health Section of the Institute for Community Medicine (ICM) at University Medicine Greifswald is analytical epidemiology and risk factor research, epidemiology of and research into health care, health systems and transfer research, disease prevention and the promotion of health, and the development and practical implementation of new models in all areas of medicinal care.

In order to complete the processing of large amounts of data (particularly personal data) that is required, competencies in the field of medical informatics and data management have been significantly expanded in recent years. The Institute for Community Medicine has comprehensive experience in the integration of complex clinical and epidemiological data from multicentre studies and decentralised field surveys in central management units based on central database systems.

1.3 Department of Medical Informatics, University Medical Centre Göttingen

The Department of Medical Informatics was founded in 1972 and since 1999 has been dedicated to collaborative medical research. It develops new methodical processes for interdisciplinary multicentre collaboration in cooperation with the researchers from Germany and abroad. The institute has provided and continues to provide methodological and technical support for many individual studies and the competence networks for dementia, congenital heart defects and multiple sclerosis, as well as for several clinical research areas and the German Collaborative Research Centre 1002 (Sonderforschungsbereich 1002 [SFB 1002]). Furthermore, the Department of Medical Informatics has contributed to the establishment of the research body known as Technology, Methods, and Infrastructure for Networked Medical Research (Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V. [TMF]). Additionally, the institute was Germany's main representative for grid-based research projects in medicine.

Currently almost 40 scientists, documentation officers and administrative staff are working in close cooperation with other institutions on the aforementioned long-term projects and on methodological support for the German Centres for Health Research (DZGs). While the competence networks in medicine represent the decade of methodological breakthrough in collaborative medical research, the infrastructure for the national centres is being established on a new level in coordination with the TMF. This new level bundles the experiences of the last ten years and combines them with the data and requirements of personalised medicine. Since around 2014, the participating associations have been able to rely on the use of a powerful research infrastructure that stands up to international comparison.

Central Data Management joint project

With its communication dated 26 March 2013, the DZHK Board of Directors commissioned University Medicine Greifswald (represented by the ICM) and the University Medical Centre Göttingen (represented by the Department of Medical Informatics [MI]) to implement the Central Data Management (CDM) unit as a joint project for multicentre studies and registers within the DZHK. This federal approach presents the opportunity to realise spacial and organisational separation of medical and personally identifiable information in accordance with a uniform standard. Medical research data is usually captured as part of multicentre studies. The following studies and registers have been selected for funding as part of an external assessment process, the requirements of which have been included insofar as they were known at the time of the creation of this document:

1. Early Versus Late Left Ventricular Assist Device Implantation (VAD)
2. Systolic Dysfunction to Congestive Heart Failure Cohort Study (TransitionCHF)
3. Translational Registry for Cardiomyopathies (TORCH)

The Central Data Management unit is a joint project between partners that are equal and operate under their own authority. It consists of the Trusted Third Party (TTP) (under the responsibility of the Institute for Community Medicine at University Medicine Greifswald [Prof. Dr. med. Wolfgang Hoffmann]), the Data Handling (DH) unit (under the responsibility of the Department of Medical Informatics at the University Medical Centre Göttingen [Prof. Dr. med. Otto Rienhoff]) and the IT Lab situated at the DZHK Main Office (Prof. Matthias Nauck).

The responsibility for ethical aspects and their harmonisation in the DZHK is realised by the Institute of Epidemiology II at the Helmholtz Zentrum (Prof. H.-E. Wichmann) and the Department of Medical Statistics and Epidemiology (Prof. Frank Kuhn) at the Technical University of Munich as part of an additional ethics project by DZHK Programme Group 7. Furthermore, every study and register independently creates a specific ethics concept and an informed consent (IC) form, which they submit for assessment by the responsible ethics committees and then send to the persons responsible for ethics at the CDM unit after a positive ethics vote. The task of the individual study centres is to recruit participants and capture data afterwards. An IT Office (comprising an IT Lab and the DZHK Main Office's IT Management Office) is to be established at the DZHK Main Office, which is based in Berlin. Among others, this office will assume the function of the DZHK Commissioner for Data Protection and responsibility for coordination with the relevant federal state authorities.

The relationships mentioned are illustrated in the organisation chart below (see Figure 1).

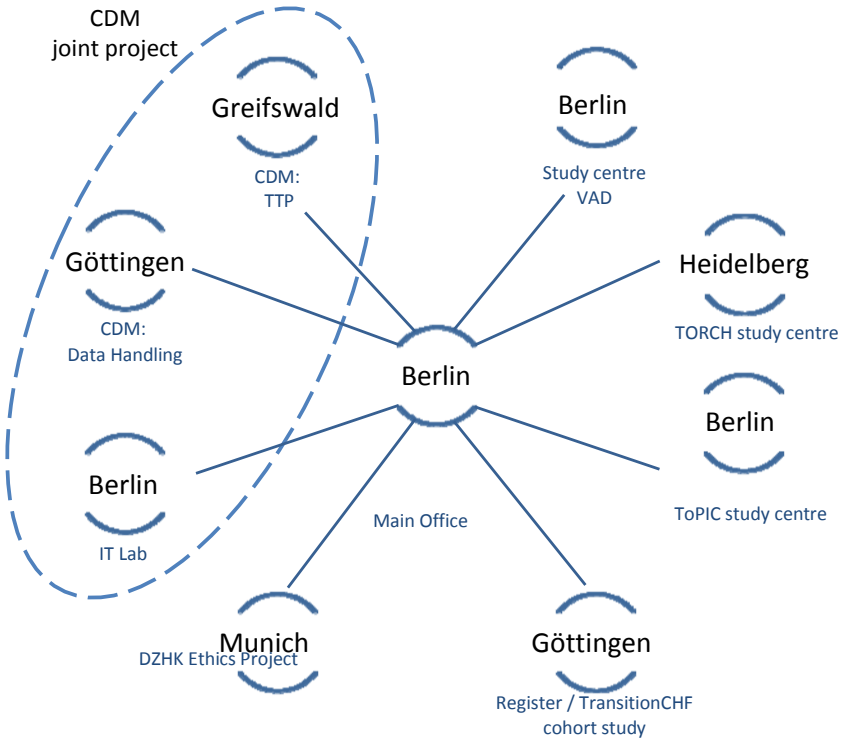


Figure 1: DZHK Central Data Management joint project partners

1.4 Involved institutions

The Ethics Project is responsible for preparing the required informed assent and consent forms in terms of content. The CDM joint project uses only the informed consent forms that have been agreed upon and harmonised in advance.

The services provided by, responsibilities of and relationship between the partners involved are regulated by separate cooperation agreements. These are currently still being coordinated between the partners and will ensure that all partners can operate equally, independently and under their own authority. Plans include both an internal cooperation agreement for the CDM joint project and a cooperation agreement between the CDM unit and the Ethics Project. Furthermore, cooperation agreements between the individual projects (registers/studies), the CDM unit and the Ethics Project are in the process of coordination.

1.5 Tasks

The technical and organisational measures necessary to successfully implement the Central Data Management unit can essentially be reduced to the three following aspects. These are illustrated additionally in Figure 2.

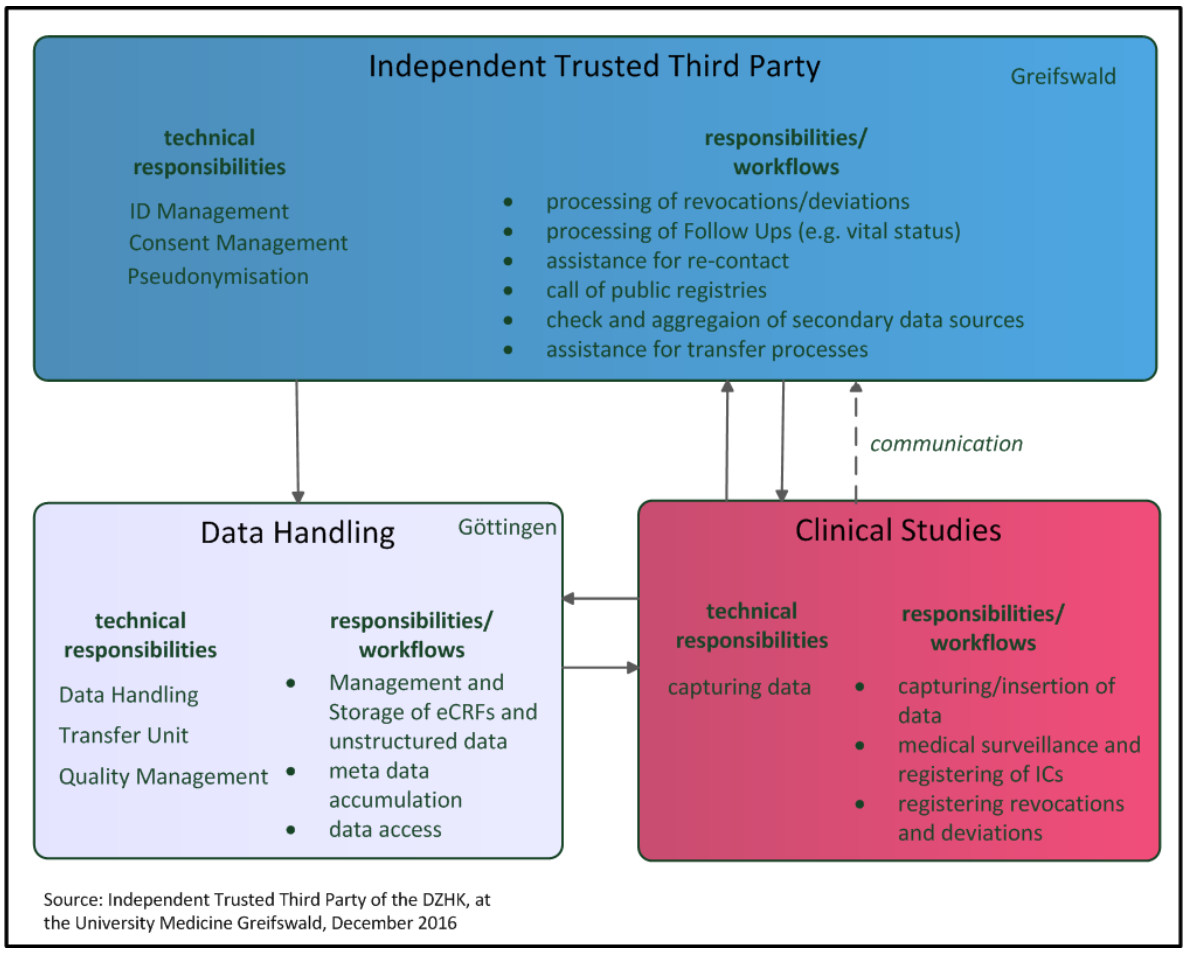


Figure 2: Responsibilities within the Central Data Management unit

Trusted Third Party (TTP)

The Trusted Third Party assumes responsibility for the management and storage of the personally identifiable information (IDAT) as part of the transfer of functions. The transfer of data to the TTP is coordinated by the participating studies, registers and cohorts (SRC) with the respective German federal state commissioner for data protection. Data processing within the TTP must be based on informed consent without exception.

Three essential technical functionalities are necessary for the management of the personally identifiable information: a master person index (MPI), consent management (CM) system and a generalised pseudonymisation service (PSN). The trusted third party receives the personally identifiable information and creates a suitable number of pseudonyms, which are then used and stored with the medical data at the Data Handling unit (see Section B1).

Data Handling (DH) unit

The data handling for all SRCs conducted within the DZHK takes place in Göttingen. This means that a study database is established and operated, and that the data elements for the individual SRCs are modelled there. To achieve this, the personnel in Göttingen are in close coordination with the chief investigators. This coordination is partly advisory (i.e. assistance is provided for uniform, sustainable selection of items for documentation) and partly practical (implementation). Trained documentation officers and documentation officers with an additional master's degree in Medical Informatics are available for these advisory and implementation services. Furthermore, a metadata directory is to be built in the DH in order to fulfil two additional tasks for the long-term comparability of DZHK studies: The first task is to allow a specific item (e.g. diastolic blood pressure) to be captured equally (e.g. sitting after five minutes' rest), stored (unadjusted two- to three-digit integer values) and provided with the same name within the database for every study conducted (e.g. bld_prsr_dia). This allows long-term comparability of the DZHK studies conducted. The second task is to allow specific information to be stored within a metadata directory, e.g. information on the circumstances (environmental metadata) of an examination. The same will apply to other data types later on, such as images or information on collected biomaterials.

Other (sometimes complex) data types will become relevant for the DZHK and CDM unit both with the three approved DZHK studies and with new projects in the coming years. This could include electrophysiological data, image data (in coordination with the corresponding DZHK working groups), follow-up data generated from a mobile source (possibly by the patients themselves) or the aforementioned examination metadata. During the course of the project applied for and together with the IT Management Office at the Main Office and TTP employees, DH unit employees create concepts to connect additional databases such as the central DZHK biobank or a DZHK image database and then realise these concepts. These databases can be operated in Göttingen, at other DZHK sites, or with external service providers. However, the CDM unit will ensure uniform architecture for the DZHK in each case. This will be achieved by strategy coordination, implementation planning (and, if necessary, realisation) as well as by cooperation on the specification of processes for establishment and operation.

Data exchange and interfaces

Some of the applications require data exchange between the TTP and the DH unit, or between the TTP and the studies. This can take place either automatically or manually using either electronic means or paper documents. Common interfaces and processes that meet the requirements of data protection law both with regard to electronic exchange and (partially) manual workflows are defined not only by the spacial separation of the data, but also by the different technological systems. The following chapters explain assumptions, conditions and applications for the TTP during interaction with users (DH unit or studies), identify interfaces and define these in greater detail.

A2 Specifications for the TTP subproject (Greifswald)

1 Workflows and data flows

The following figures describe the status of the planning and consultation at the time the document was created.[1]

1.1 Requirements and technical approach

In addition to the aspects specific to the TTP that are described here, detailed coordination of all processes with regard to a common interface is necessary for both the DZHK and the Data Handling subproject. This coordination is to be achieved with a focus on technical characteristics and processes (workflows and use cases). It is also necessary to define and describe the internal processes of the TTP.

Cooperation between the TTP and the University Medical Centre Göttingen's Data Handling subproject means that the following conditions are necessary for the implementation of the TTP:

Pseudonymised medical data must be captured in the study centres using electronic case report forms (eCRFs). A basic dataset must be used for all studies, registers and cohorts. Each study must also have a study-specific dataset. IDAT must be captured in order to create the required pseudonyms. However, this is not to be stored in the eCRF.

In the medium- to long-term, unstructured data (e.g. image data, ECG data or free text questionnaire data) will be captured. This (MDAT) data may also be processed by the DH unit in Göttingen or by other institutions. These institutions must provide the same required technical infrastructure. The required transfer office (for the transfer of the collected data for research purposes) should be made available by the DH unit. This data protection concept must be expanded if further data types and/or institutions are added.

The eCRFs are developed with the help of the web-based tool secuTrial, which was developed by iAS Berlin GmbH. The sole contract partner of iAS GmbH is the University Medical Centre Göttingen. The cooperation agreement of the CDM joint project stipulates that the TTP has a right to be involved in the definition and testing of the interface. The relevant studies are responsible for the form and content of the electronic report forms.

[1] The purpose of this status is to begin data capture. The plan is to update the content of the versions (e.g. by adding additional processes) as soon as the defining aspects have been coordinated between the partners involved. The proposed extensions are support during the transfer of data and materials (transfer office process) and the integration of modalities other than eCRFs (e.g. the biobank and imaging procedures).

secuTrial has a central roles and rights system. As required, it is therefore only possible to allow a selection of persons or groups of persons (doctors, studies) to amend identifiable information. Defined, contractually agreed interfaces allow for the necessary interaction between the Data Handling unit, Trusted Third Party and secuTrial.

There are also plans to manage biosamples using a laboratory information and management system (LIMS). The current plan is for the DZHK Main Office's IT Management Office to prepare an invitation to tender, the specifics of which are currently being developed.

Personal data, consents and authorisations, as well as all the pseudonyms and allocations created at a single point in time, are saved at the TTP. The TTP manages both this data and other metadata on studies, projects and locations. The data is encrypted and stored in accordance with the TTP's data protection concept and the relevant security regulations with regard to access and backup (see Section B1, Chapter 4).

1.2 Data flows within the CDM

Figure 3 (below) illustrates the planned data flow between the individual subprojects.[2] The technical details for the individual processes can be found in the annex (see Figures 16-19).

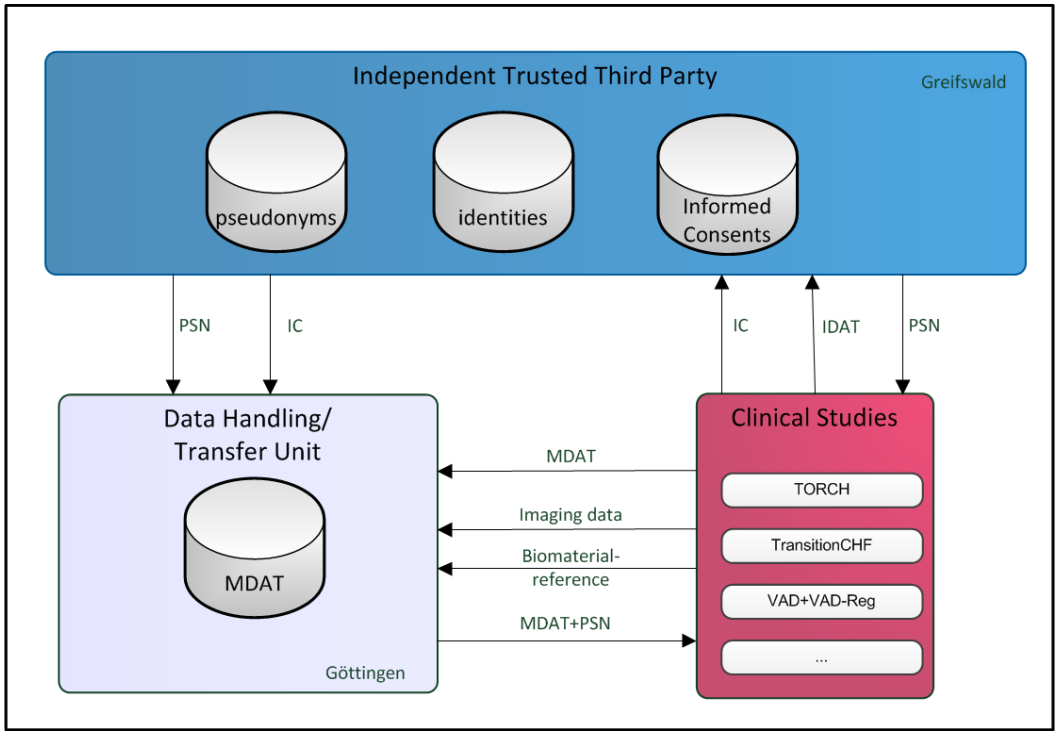


Figure 3: Data flows within the Central Data Management unit

Upon inclusion of a person, the first step is to send the informed consent information and identifiable information to the TTP. The unique PSN generated for this person by the TTP is then transferred to the study centre and (together with the informed consent information) to the Data Handling unit.

[2] No attempt will be made to illustrate the authentication mechanisms and provision of the eCRF using secuTrial at this point.

Within the framework of the DZHK, primary pseudonyms or first-level pseudonyms are at no point in time transferred to the DH unit or projects; instead the DZHK-specific TTP workflow stipulates that secondary pseudonyms or second-level pseudonyms are to be derived and transferred immediately. The DH unit stores the eCRF, laboratory / biosample and image data in various systems and uses specific secondary pseudonyms for a person. The DZHK-specific TTP workflow ensures that several unique secondary pseudonyms can be derived for a person's primary pseudonym.

Entering the PSN allows the study centre to transfer the subject's MDAT (in the form of image data and biosamples) to the Data Handling unit in Göttingen, for example.

Figure 4 additionally illustrates that transfer of personally identifiable information takes place directly between the study centre and the Trusted Third Party through a tunnel. The Data Handling unit has at no point in time access to the participant's IDAT. The TTP has at no point in time access to the participant's medical data.

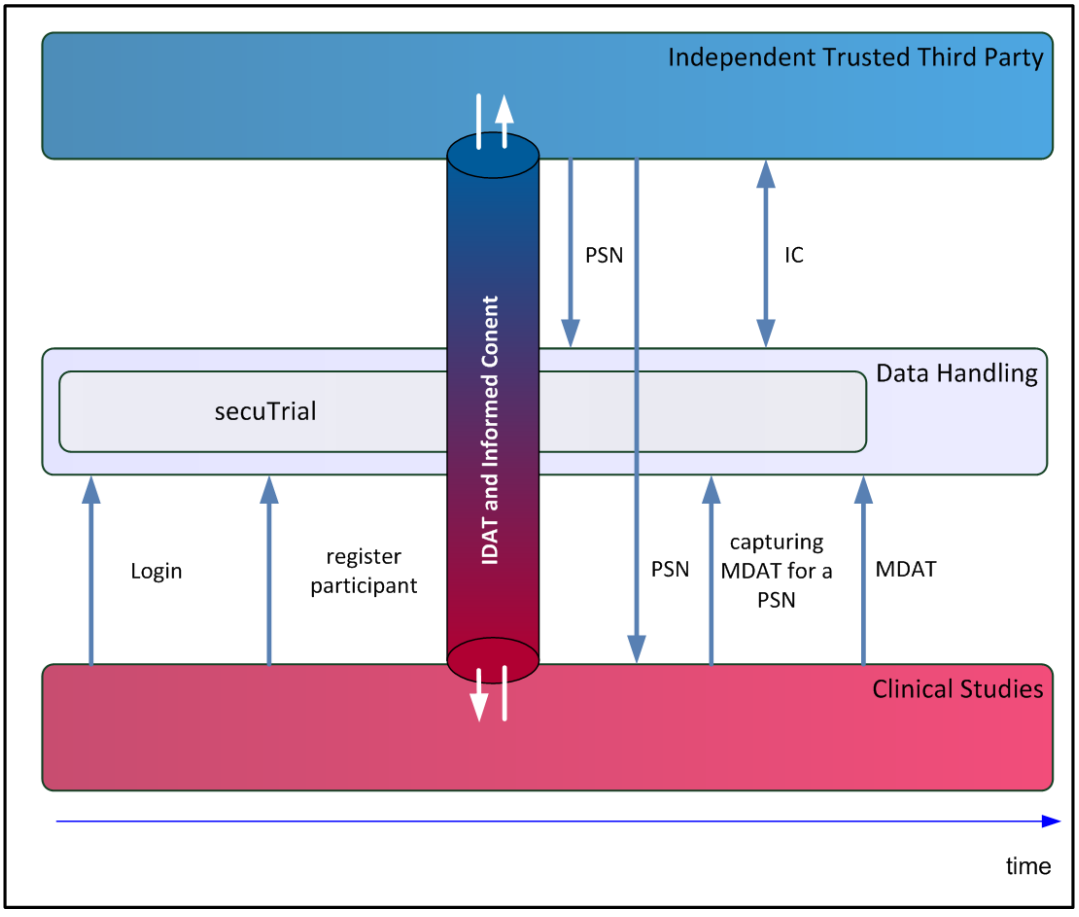


Figure 4: Interactions between the involved systems for "Creating a new participant"

Table 1 illustrates that in no Central Data Management unit subproject does IDAT, MDAT and PSN information coexist outside of the hospitals that are treating the patients and collecting the data. It is also evident that only pseudonymised data is stored within the Data Handling unit.

Subproject	IDAT	MDAT in accordance with consent	PSN
Studies (hospitals)	Yes	Yes	Yes
TTP	Yes	No	Yes
Data Handling unit	No	Yes	Yes

Table 1: Data distribution matrix

1.3 Workflows

The following describes from an operational point of view the intended applications. These clarify the cooperation of the independent Trusted Third Party (TTP) in Greifswald, the Data Handling unit in Göttingen and the study centres in the DZHK projects. The technical details can be found in the annex (Figures 16-19).

For all applications, the study-centre-based user[3] must first log in to the study-specific study portal / eCRF system. This authentication assigns role-based rights to the user.

Logging in to the eCRF system

User authentication via secuTrial is necessary in order to be able to use the functions of the Trusted Third Party in the study centre. secuTrial manages the user roles and rights (trusted delegation) on the behalf of the Trusted Third Party. This allows the Trusted Third Party to offer a role-specific functionality. If authentication does not take place, it is not possible to use the Trusted Third Party functionality. The same is valid for attempted use of Trusted Third Party functionalities without using the secuTrial system for the Data Handling unit in Göttingen.

The login process is initiated by the study-centre-based user. The mask for entering the login credentials is provided by the Data Handling unit in Göttingen. The user enters a user name and password, and the data is sent to secuTrial in encrypted form via HTTPS (see Chapter 4.4). After the login credentials have been verified and the roles and rights assigned, session-based communication with the Trusted Third Party is established – also via HTTPS. The login information sent (user name, institution, user ID, role, forename, surname, title) also serves as audit information within the Trusted Third Party (see Chapter 4.3). Session-based communication between secuTrial and the study-centre-based user is then established so that queries can be directly forwarded to the Trusted Third Party. Figure 5 illustrates the relationships described.

[3] A user is a natural person who creates, amends or works with data in the context of the project.

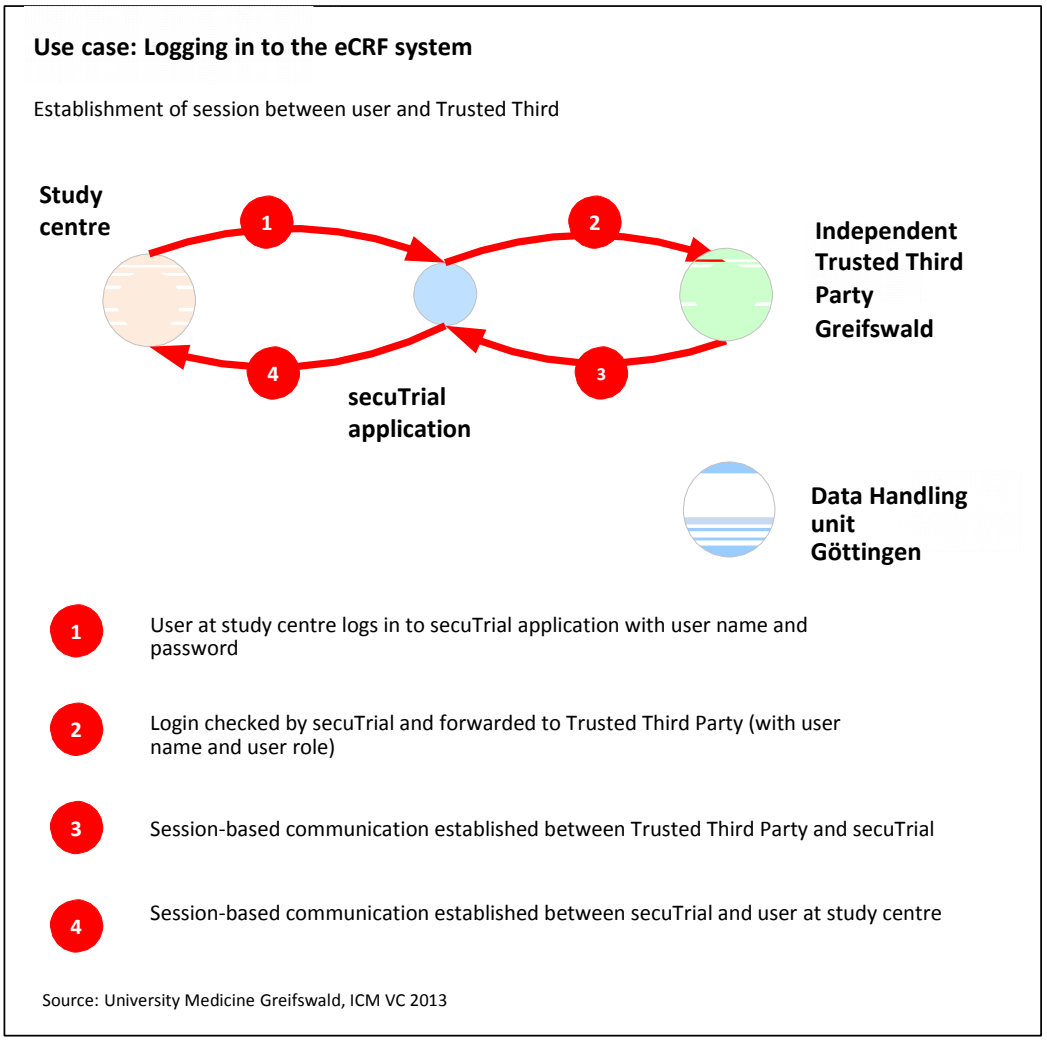


Figure 5: Logging in to the eCRF system

Creating study participants and informed consent

To create another person, the study-centre-based user calls up the relevant eCRF. The form is provided digitally by the Trusted Third Party. The user enters the identifiable information in accordance with the specifications of the Trusted Third Party (see Section B1, Chapter 3.2) into the input mask and sends it directly to the Trusted Third Party via an encrypted tunnel (see Chapter 1.2 and Figure 7).

Similarly, the Trusted Third Party provides an eCRF for entering consent in order to allow transfer of the paper IC forms at the study centre to the Trusted Third Party as a digital representation (see Chapter 3.3). The informed consent status can be positive (i.e. at least partial consent is given) or negative (i.e. consent is not given). After the informed consent status has been entered it is sent to the Trusted Third Party as well.

Within the Trusted Third Party, matching processes are initiated based on the IDAT received. These processes compare the IDAT received with the IDAT that is already known. If the specified person does not yet exist and the IC is positive, a new person is created and a corresponding pseudonym is generated at the TTP (see Section B1, Chapter 2.2). This pseudonym is sent back to the study centre, which can then begin capturing the medical data using the pseudonym. The necessary eCRFs are made available by the Data Handling unit in Göttingen, which then receives the pseudonymised data. Figure 9 illustrates this (ideal) scenario from an operational point of view.

If during the matching processes it emerges that the identified person already exists in the Trusted Third Party's system and the IC is positive, the study centre is informed to this effect and the Trusted Third Party returns the already assigned pseudonym. If there is a critical match during the matching process, i.e. the IDAT provided corresponds to an existing person with only a few differences, the study-centre-based user is requested to check and if necessary correct the data entered to ensure the information is correct. If there is another critical match after renewed matching, i.e. double entries cannot be entirely ruled out (see Section B1, Chapter 1.1), a new person is created using the IDAT sent provided there is a positive IC. Further steps follow as in the ideal case described above.

If participation is refused by the patient or subject (negative IC), the remaining process is based on the design and content of the informed consent forms (see Chapter 3.3) and contains two possible approaches. If there is no positive IC, the first of these is, as legally stipulated, to store neither the IDAT nor the IC form itself at the Trusted Third Party. The second is to store the data if the participant agrees in advance to storing the IDAT and IC form even if the IC is negative.

Both options are possible without any technical difficulties. Which option is actually realised depends on the manner of the explanation and the stipulations of the Ethics Project (e.g. if a minimum IDAT dataset is to be stored as a matter of principle).

The current stipulations for the DZHK are that neither the IDAT nor the IC form itself should be stored at the Trusted Third Party (see Chapter 3.3). In such a case, no new person or IC is created.

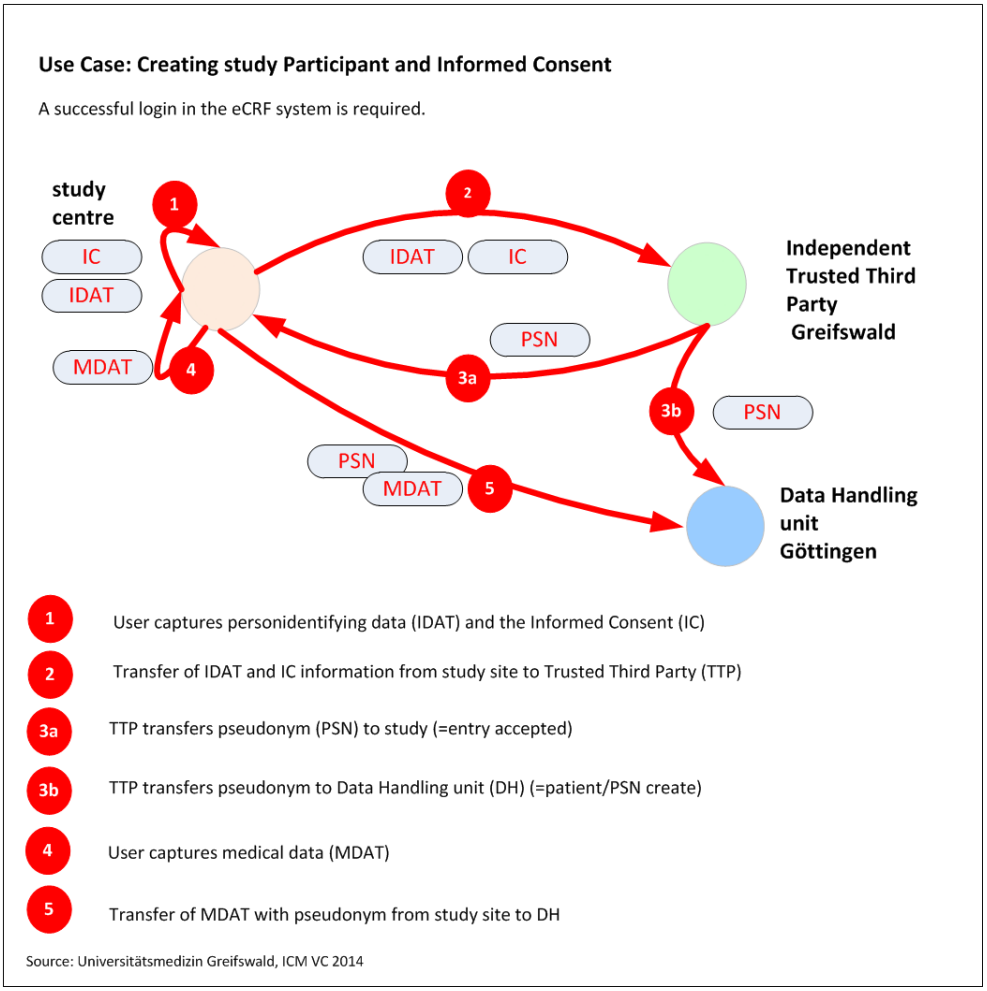


Figure 6: Creating a new study participant

Updating a study participant

There are essentially two types of changes to personal data. If a person's IDAT is to be updated (e.g. change of name), the necessary change to the dataset is made at the Trusted Third Party. If, on the other hand, changes or additions are to be made to MDAT, the changes are made at the Data Handling unit. The cooperation agreement stipulates that the eCRFs necessary for changes are to be provided via the unit that saves the data.

The prerequisite for updating a person is that the person has already been created and that the corresponding pseudonym thus is known. The study-centre-based user calls up the corresponding eCRF in secuTrial by entering the pseudonym of the person to be updated. Different forms are available for editing IDAT and MDAT.

To update IDAT, the user enters the data to be updated in the eCRF and sends it to the Trusted Third Party. Before the updated IDAT can be saved, it must complete the matching process. If the new IDAT is unique, the dataset is updated with the same pseudonym at the Trusted Third Party.

If the matching process results in a critical match, the study-centre-based user is informed and further steps must be manually cleared with the responsible data custodian at the Trusted Third Party.

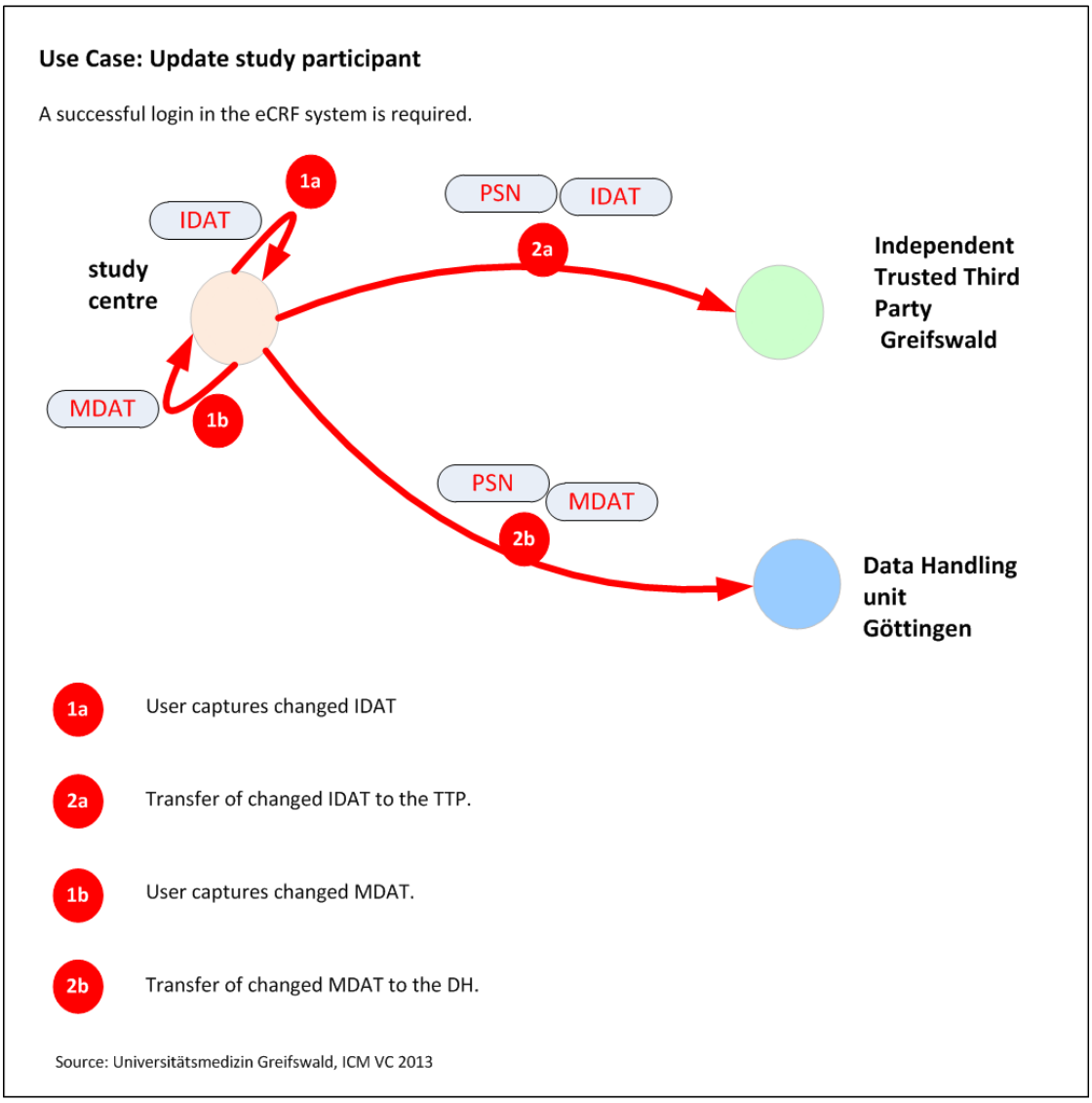


Figure 7: Updating a study participant

Changes to MDAT are analogue. The pseudonym is used to request the relevant eCRF from the Data Handling unit, to update the medical data and to transfer the data for storage at the Data Handling unit in Göttingen. Figure 7 illustrates the workflow described.

Updating informed consent

The prerequisites for updating informed consent status are that the associated person must already have been created, the corresponding pseudonym must already been known and the person must have already been assigned an IC. The study-centre-based user calls up the input mask for creating an informed consent entry in secuTrial using the relevant person's pseudonym. This form is made available by the Trusted Third Party.

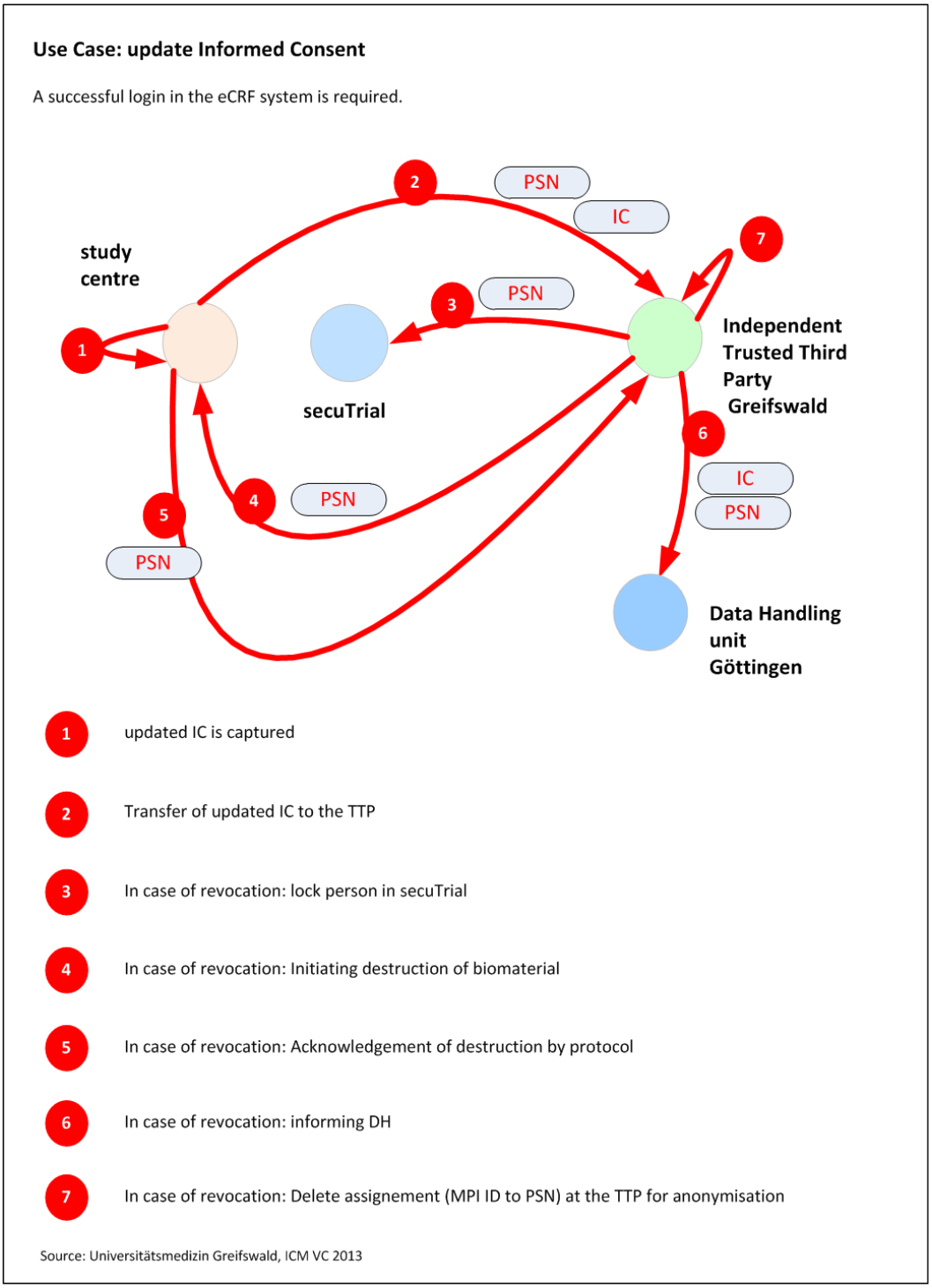


Figure 8: Updating the informed consent status of a study participant

Updating the informed consent (IC) status essentially consists of capturing a new IC. After the IC is created, it is the basis for decisions on the admissibility of the individual data processing processes and constitutes the current consent status for the person concerned. If the newly created IC status is positive, all data collected until this point will continue to be processed (stored and used) under the conditions at which consent was given. Future data capture, however, is based on the conditions of the new IC. A negative IC constitutes a revocation.

The new IC is transferred to the Trusted Third Party together with the pseudonym. In the event of a revocation, the TTP arranges all further steps (see Chapter 3.3). The study-centre-based user is then informed that the IC has been successfully updated. Figure 8 illustrates the relationships described.

2 Further protection requirements

Within the DZHK Central Data Management unit, the Trusted Third Party subproject has no further protection requirements with regard to the illustration in Section B1, Chapter 3.3.

3 Supplementary technical and organisational measures

3.1 Integrating the secuTrial eCRF system

The Data Handling subproject uses electronic case report forms (eCRFs) to capture the data. The commercial eCRF solution secuTrial is used to do this. secuTrial is entirely web-based and supports typical functions such as audit trail, a roles and rights concept, and electronic signature. There is official confirmation of compliance with the typical security standards according to 21 CFR Part 114. [2] A detailed description of the data protection aspects, which goes beyond the interface use illustrated here, can be found in the DH unit's data protection concept.

3.2 Separating identifiable information

Using secuTrial allows the earliest possible secure separation of medical and personally identifiable information from an organisational point of view.

The study participant's IDAT is replaced in advance by a unique pseudonym within the DZHK environment. This pseudonym is assigned by the Trusted Third Party. It is not possible to trace the pseudonym to the IDAT without the involvement of the Trusted Third Party.[5]

The Trusted Third Party uses two systems. The central ID management system supplies the unique identification for persons using the IDAT and assigns an individual key to each person. This key is made indecipherable using a separate pseudonymisation process. Both the person code (MPI ID) and its assigned pseudonym (PSN) are electronically managed and stored within the technical infrastructure of the Trusted Third Party.

The advantages of this process are that medical data can be captured in pseudonymised form, medical data cannot at any time be processed within the Trusted Third Party, and that identifiable information is not at any time present at the Data Handling unit.

[4] http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfCFR/CFRSearch.cfm?CFRPart=11

[5] At least not without a disproportionately high level of effort (see German Federal Data Protection Act [BDSG] § 3 paragraph 6a)

3.3 Dealing with informed consent forms

The consent of the person to participate in the respective study or register is documented in an informed consent form. The participant legally confirms (usually by signature) that s/he agrees to the capture, processing and storage of his data to the extent described and authorises the responsible persons within the DZHK to collect his data for research purposes. The informed consent information is converted into digital format (a version that can be used electronically and a PDF version) by entering the information into a corresponding input mask and by scanning the paper informed consent form as an image.[6] The original paper document remains in the study centre for the duration of the study. In addition to the electronic capture of the IC using the eCRF, the scan of the original paper document is sent to the TTP as an image. The reason for this is traceability in the event of an audit or quality assurance.

Doctors inform potential participants on site at the study centres of the extent and consequences of their consent and the opportunities to revoke their consent. This process ensures that the participant has understood the individual items of the respective IC form and is aware of their implications. The participant has the opportunity to ask questions at any time. These questions are answered immediately by the doctor obtaining consent.

The IC form used is modular. This allows the participant to consent only to selected sections. Although the storage term for the data is primarily legally stipulated, it is also based on the text of the IC form.

Participants have the right to revoke their consent completely at any time with or without providing a reason (see Section B1, Chapter 1.3). If there is no revocation, the use of the collected research data is based on the IC form that was relevant when the data was collected. It is not possible to cancel the revocation.

The revocation can be presented to the persons responsible for the study in written form (letter, fax or email) or orally. The persons responsible for the study then communicate the revocation to the Trusted Third Party. The revocation is applicable to all the data collected to date. The Trusted Third Party is responsible for the documentation and implementation of the revocation. The Trusted Third Party continues to store both the IDAT and revocation of the participant. In the case of a revocation, the collected data is anonymised by deleting the corresponding assignment at the Trusted Third Party (the captured data is not deleted[7]). The Trusted Third Party also initiates the destruction of biosamples for the unit that has the physical samples. Confirmation of the successful enforcement of the revocation is communicated to the relevant study centre in writing.

[6] The precise design and shape of the informed consent forms for DZHK projects is not yet known at the time of the creation of this documentation and will be developed and harmonised during the course of the Ethics Project's work. At present, it is assumed that the master version of the IC form will be presented to each participant on paper.

[7] Rather than being deleted, the data is locked in the primary database. Both this and reviews of up-to-date IC statuses before the publication of the data by the transfer office ensure that the data will no longer be used.

3.4 Personnel measures

In addition to Section B1, Chapter 4.8, the Trusted Third Party employees consist of the Coordinator of the Trusted Third Party and two research associates for organisational processes and their technical implementation. The group of Trusted Third Party employees is financed using project funds and DZHK employees have been informed of the group members' names.

A3 Specifications for the Data Handling subproject (Göttingen)

1 Workflows and data flows

The data located at the Central Data Management unit's Data Handling (DH) unit, which is based at the Department of Medical Informatics at the University Medical Centre Göttingen, is tasked with developing and providing methods, processes and tools for processing[8] personally identifiable information[9]. This task can essentially be divided into two main processes: Data input via electronic case report forms (eCRF), data export and release of this export for subsequent analysis.

Other related tasks, such as expert advice during conceptualisation and creation of the eCRF, training and support for the software used (secuTrial), and the establishment of a reporting component for quality assurance concerns on the part of the Management Board and the Main Office. According to the generic data protection concepts published by the TMF, the Data Handling unit is responsible for processing the pseudonymised medical data (MDAT) and operating the research database [3]. Identifiable information (IDAT) is at no time known to the software systems or DH unit employees.

The typical tasks of the Data Handling unit include:

- Preparing and operating the secuTrial study database on suitable servers
- Implementing any eCRFs (base dataset and study-specific modules) required for the data collection in the DZHK
- Processing the data captured via secuTrial using the pseudonyms generated by the Trusted Third Party

The following process description deals only with the secuTrial tool by interActive Systems GmbH[10], which has already been put in place and is used to capture, process and store study data. The present document must therefore be appropriately expanded if other capture systems (e.g. a biomaterial management database or image data management system) are added. The DZHK intends to obtain and operate central storage and management systems for *image data* and *biomaterial management data*. Data transfer between these software components is sensible (e.g. it is conceivable that clinical data and the information "Smoker: yes/no" are displayed directly within the image analysis software during evaluation of an image dataset of the heart).

---

8 Data processing is used here in accordance with the definition in the Landesdatenschutzgesetz Niedersachsen (Data Protection Act for the German Federal State of Lower Saxony [NDSG]) § 3 paragraph 2 and comprises the collection, storage, amendment, transfer, locking, deletion and use of personal data [13].

9 "Personally identifiable" is used here in accordance with the definition in the notes on the NDSG. The person can be determined if the person can be identified with the help of additional knowledge from the unit responsible for processing the data [13]. This additional knowledge is managed by the Trusted Third Party.

10 http://www.secutrial.com

Although this is possible through communication between secuTrial and the image data management software, it has not yet been taken into account in this data protection concept.

1.1 Data collection

Data collection takes place in the centres responsible for the respective studies. These centres are named in both the respective funding applications and the study designs. Because of their connection to the treatment process, the centres that collect the data have the right to information with regard to identifiable information and medical data (see Table 1 p. [15]). The person collecting the data has direct patient contact. The data is usually collected by persons specially trained for the studies (e.g. a study nurse or study doctor). The staff responsible for collecting the data ask the patient for their consent to store, amend and use the data to be collected in accordance with NDSG § 10 paragraph 2 item 1 and make a record of this consent. This consent information is stored at the Trusted Third Party (see the use case entitled "Creating study participants and informed consent" (p. 16) and the use case entitled "Updating informed consent" (p. 19).
„Teilnehmer und Informed Consent anlegen", Seite 16 sowie Use Case „Informed Consent aktualisieren", Seite 19).

1.2 Data capture

Data capture using secuTrial software is achieved by entering the items queried during collection using the eCRFs created and provided by the DH unit. Although the data collection process runs parallel to the data capture process in most cases, it is not absolutely necessary from a technological point of view . The collected data can initially be stored intermediately on printed questionnaires. The collected data is captured in the eCRFs provided and is usually captured by a study nurse or study doctor. Access to the secuTrial study database and its technical protection is described in Section 8.1 (p. 52 ff.).

The MDAT and IDAT are separated before data collection. This is achieved by incorporating an IDAT input mask provided by the TTP before the MDAT is actually entered. The IDAT is entered directly on the TTP servers. This tunnelling ensures that the IDAT is at no time known to the secuTrial system. This is illustrated in Figure 4 (p. 14), Section 3.2 (p. 21 ff.) and Section 2.4 (p. 36 ff.).

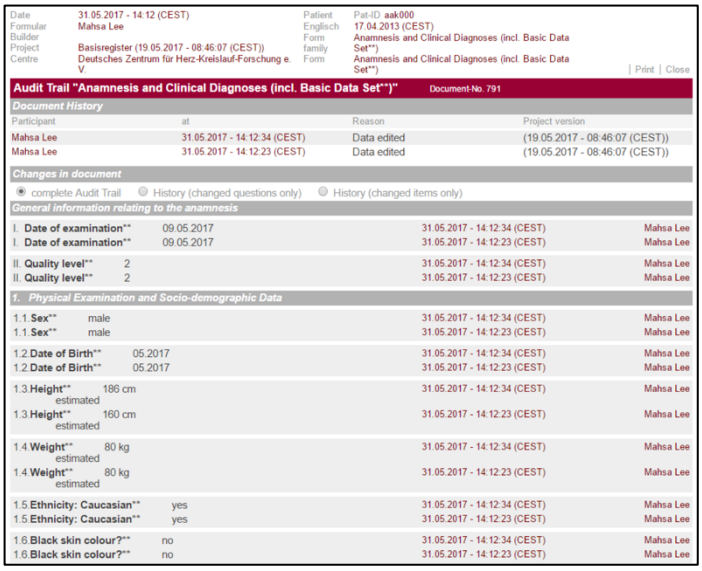
secuTrial also implements query management within studies, registers and cohorts. Queries are designed to help monitors and other authorised users to investigate unclear entries, etc. Study doctors and study nurses can read, review, answer (if necessary) and close queries. This process can also only be accessed by users who are set up in the system.

1.3 Data storage and management

In addition to providing suitable capture tools, another core task of the Central Data Management unit is the long-term storage and management of the collected data. A non-bypassable audit trail is created for the corresponding dataset as early as the data capture stage.

This audit trail saves which person (login information in secuTrial) with which role (derived from the login information) conducted which operation at which time / on which date and on which dataset. This allows consistent version control and traceability for all amendments. The aforementioned query management is also recorded in its entirety in the audit trail. The audit trail offers an overview of all changes made to the data and saved in the up-to-date form. It can be accessed after the form has been saved for the first time.

Entering and saving comments, conducting and answering queries, conducting SDV[11], reviews and form-locking actions, and ending data capture are all storage processes for the respective form. For this reason, all these actions are illustrated in the storage history in the upper part of the audit trail. Every storage operation documents the current project version so that changes to the project setup can be traced here as well. It also displays whether an e-signature was used for saving and if it is still valid.


Date 31.05.2017 - 14:12 (CEST)
Formular Builder Mahsa Lee
Project Basisregister (19.05.2017 - 08:46:07 (CEST))
Centre Deutsches Zentrum für Herz-Kreislauf-Forschung e. V.
Patient Pat-ID aak000
Englisch 17.04.2013 (CEST)
Form family Anamnesis and Clinical Diagnoses (incl. Basic Data Set**)
Form Anamnesis and Clinical Diagnoses (incl. Basic Data Set**)
Print | Close

Audit Trail "Anamnesis and Clinical Diagnoses (incl. Basic Data Set**)" Document-No. 791

Document History

Participant	at	Reason	Project version
Mahsa Lee	31.05.2017 - 14:12:34 (CEST)	Data edited	(19.05.2017 - 08:46:07 (CEST))
Mahsa Lee	31.05.2017 - 14:12:23 (CEST)	Data edited	(19.05.2017 - 08:46:07 (CEST))

Changes in document

complete Audit Trail | History (changed questions only) | History (changed items only)

General information relating to the anamnesis

I. Date of examination**	09.05.2017	31.05.2017 - 14:12:34 (CEST)	Mahsa Lee
I. Date of examination**	09.05.2017	31.05.2017 - 14:12:23 (CEST)	Mahsa Lee
II. Quality level**	2	31.05.2017 - 14:12:34 (CEST)	Mahsa Lee
II. Quality level**	2	31.05.2017 - 14:12:23 (CEST)	Mahsa Lee
1. Physical Examination and Socio-demographic Data			
1.1. Sex**	male	31.05.2017 - 14:12:34 (CEST)	Mahsa Lee
1.1. Sex**	male	31.05.2017 - 14:12:23 (CEST)	Mahsa Lee
1.2. Date of Birth**	05.2017	31.05.2017 - 14:12:34 (CEST)	Mahsa Lee
1.2. Date of Birth**	05.2017	31.05.2017 - 14:12:23 (CEST)	Mahsa Lee
1.3. Height**	186 cm estimated	31.05.2017 - 14:12:34 (CEST)	Mahsa Lee
1.3. Height**	160 cm estimated	31.05.2017 - 14:12:23 (CEST)	Mahsa Lee
1.4. Weight**	80 kg estimated	31.05.2017 - 14:12:34 (CEST)	Mahsa Lee
1.4. Weight**	80 kg estimated	31.05.2017 - 14:12:23 (CEST)	Mahsa Lee
1.5. Ethnicity: Caucasian**	yes	31.05.2017 - 14:12:34 (CEST)	Mahsa Lee
1.5. Ethnicity: Caucasian**	yes	31.05.2017 - 14:12:23 (CEST)	Mahsa Lee
1.6. Black skin colour?**	no	31.05.2017 - 14:12:34 (CEST)	Mahsa Lee
1.6. Black skin colour?**	no	31.05.2017 - 14:12:23 (CEST)	Mahsa Lee

Figure 9: Example of an audit trail within secuTrial All entries, amendments and deletions are indelibly logged.

1.4 Transfer Office: Data preparation and transfer

The secuTrial system operated by the DH unit is primarily for documenting and storing study data. Although filtered exports of this database are also possible, the system itself is not designed for conducting statistical analyses or for complete quality management of the data. Downstream systems, which also operate exclusively with pseudonymised data, are used for this purpose. The data is provided by the Transfer Office, which is also developed and operated by the DH unit. Data preparation involves the following steps:

- Exporting the data from the study database

[11] Source data verification (SDV) can be configured as an additional data quality assurance step. Comparisons between the study data and the original data (see the steps entitled "Data collection" and "Data capture") can then be logged in secuTrial.
It is essentially possible to check each item, but summary or random checks are also possible for forms, visits and all patients.

- Replacing the pseudonyms used in secuTrial (PID_MDAT) with a pseudonym generated by the TTP (PID_DWH)
- Loading these re-pseudonymised data onto the DZHK Transfer Office's data storage unit

Specific search queries can be made with regard to the pseudonymised data within the Transfer Office data storage unit.
These search queries are processed by the Transfer Office in accordance with the DZHK e.V. Use and Access Policy, i.e. all queries are stored - even if they are later discarded. The data published by the Transfer Office is also stored with its specific information (which data was transferred to which person when). The released data is not sent to the querying person using the pseudonym used by the DH unit; instead the data is re-pseudonymised. The released pseudonym is also generated at the TTP.

Using the data for research

Once data preparation and data transfer preparation begins, data use is immediately divided into two scenarios. The first usage scenario is the use of the research data by researchers associated with the DZHK. This is the primary usage scenario and requires that the aforementioned DZHK Use & Access Committee is notified of and approves the proposal for use (see Use and Access Policy § 5). Data is not released in any form to persons who are not members of either the CDM unit or the Use & Access Committee without a positive vote. First, this process is necessary to ensure a clear position on use and property rights. Second, during the export process the data is again re-pseudonymised so that the exported data receive an individual pseudonym. Thanks to this two-stage pseudonymisation, the pseudonyms provided by the custodian are restricted to CDM systems. Anonyms will not be used in this case to allow for return of potential research results to the DZHK. This process is illustrated in Figure 20 (p. 57).

Quality management and controlling

The second usage scenario in terms of preparation and transfer is the use of data for internal quality management within the DH unit and for reporting to the controlling bodies of the DZHK (the Management Board and Main Office). In contrast to the aforementioned research exports, no data is released by the CDM unit at this stage; not even in anonymised form. For this purpose, persons assigned to perform quality management can report quantitative indicators to the DH unit. These indicators are then registered there with their respective rule for generation. This registration is coordinated between the QM and DH units. This process is illustrated in Figure 21 (p. 57). Modelling queries submitted to the data-handling systems and the downstream warehouse are conducted by the DH unit so that the quantitative indicators are calculated by the CDM systems and stored in a storage unit. Depending on the type of quantitative indicator, it may be necessary to merge the stored data temporarily during the calculation process. It is important to emphasise that the use of the merged data is only a temporary measure to calculate the report data. The calculation takes place on the main memory so that the data is not to be stored on non-volatile memories. The expected quantitative indicators are largely cross-sectional information, which will not be personally identifiable. The figures provided are therefore on a level of abstraction that rules out personal identification. This process is illustrated in Figure 23 (p. 58).

Architecture of the Transfer Office

The central information system for implementing both usage scenarios is a modular data warehouse system, which (along with the data capture systems) assumes the tasks of preparing and composing the data to be exported. Data processing within the warehouse is subject to a sequential process of extracting, transforming and loading the data (ETL). Depending on the use case, this process may be repeatedly adjusted and conducted. The data is extracted and merged from various source systems using adjusted export filers. The data is re-pseudonymised early in this process. The pseudonyms used by the source systems are sent to the custodian and replaced with the warehouse domain pseudonyms. At the same time, initial quantitative indicators are calculated and integrated into the data stream. Once the information for this calculation is distributed between various source systems, the custodian will temporarily merge the datasets and then discard them once the calculation is complete. During the import, the data is cleansed of artefacts and harmonised according to quality assurance aspects. Although the datasets are stored in a person-centred scheme in the warehouse, they are arranged according to various pseudonyms. Even after having been imported into the warehouse the data remain in non-merged form.

To allow the warehouse information to be made available to the researcher or controller, it must be possible to merge the data if required. This is achieved by the Trusted Third Party performing a record linkage. This involves merging the pseudonyms for export and sending them to the custodian together with a formal description of the intended use. The custodian checks each dataset to see if it is covered by informed consent and if the dataset may be used as intended. If the answer is positive, the Trusted Third Party replaces the warehouse pseudonyms with a linkage pseudonym and returns the dataset to the Transfer Office. Depending on the usage scenario, there are then two different procedures for using the merged data. The data for the controlling bodies is prepared according to the rule for generating the registered quantitative indicators. The data is only kept merged for the time required to calculate the quantitative indicators. The results are then rendered into a suitable format (e.g. OLAP format for analytical evaluations). The transfer and recipient are documented in the warehouse's audit trail and no further traces of the merged data remain in the warehouse. In the case of export for a research project, a researcher requires a merged export. If the custodian has verified that the patient has agreed to such an export and that this consent is still valid, the warehouse data is merged using the data custodian's record linkage. Individual export pseudonyms are created for every export and will replace the warehouse and record linkage pseudonyms. This process is illustrated in Figure 22 (p. 58).

1.5 Involved groups of persons

The following groups of persons at the University Medical Centre Göttingen are involved in establishing and operating the Data Handling unit:

- IT business unit staff are responsible for smooth operation, maintenance and administration of the implemented infrastructure (software and hardware).
- Department of Medical Informatics staff are responsible for the development and implementation of the overall DH concept and its constituent parts.

None of these groups of persons has access to the identifiable information of the stored datasets or is in the position to request such access.

2 Contractual basis for cooperation between the TTP and the DH unit

The CDM can perform the tasks entrusted to it only as a whole. The TTP and the DH unit are necessary components to achieve the described goals. For these two to work together, a cooperation agreement was signed between the University Medicine Greifswald, the University Medical Centre Göttingen and the German Centre for Cardiovascular Research (DZHK). This agreement was signed by all parties and came into force on 18 February 2014.

B Project-independent section

B1 Trusted Third Party

1 Independent Trusted Third Party (TTP) processes

The Trusted Third Party is essentially a technically and organisationally independent system consisting of a custodian, a defined amount of processes and the autonomous technical services necessary for these. It assumes the tasks of "data custodian" or "independent trusted authority" as illustrated in concepts including the TMF's generic data protection concepts [3].
„Datentreuhänders" bzw. der „Vertrauensstelle", wie sie u.a. in den generischen Konzepten zum Datenschutz der TMF dargestellt werden [3].

The typical tasks of the independent Trusted Third Party include:

- Assigning personally identifiable information and corresponding IDs for source and secondary systems
- Managing consents, authorisations and revocations
- Pseudonymising and de-pseudonymising data
- Retrieving register data
- Matching and merging secondary data
- Assisting in follow-ups (e.g. vital status)
- Assisting in re-establishing contacts and in communicating incidental findings
- Implementing revocations and their procedural consequences
- Assisting in the Transfer Office's data and material transfer process

The Trusted Third Party processes are secured using established internal SOPs. The following section discusses more closely the central responsibilities of the Trusted Third Party, which are essential in order to fulfil the functions conferred upon it.

1.1 Unique identification

Medical facilities typically use unique local identifiers to clearly assign a person's medical data. However, these identifiers are only valid within their respective domain (e.g. hospital). Datasets of identifiable information on the same person can deviate from one another in various sources due to spelling mistakes or interim changes, which means that assignment errors can occur during merging of data. If data from different persons is incorrectly assigned to a single person, a **homonym error** occurs. The reverse case is referred to as a **synonym error**. The former is fatal and can only be corrected subsequently with a great deal of time and effort, while the latter can be resolved technically with the aid of additional data.

In order to be able to assign research data from several projects and studies to a single person, a project-wide ID is necessary: one to which both the personally identifiable information and the individual local ID are assigned. This process must be error-tolerant and traceable.

The task of the Trusted Third Party is to assign identity data to existing datasets in a secure manner while avoiding homonym errors, to create new datasets while avoiding synonym errors, to recognise potential duplicates, to clarify any uncertain assignment and to transfer them to certain assignment, and to update existing identity data.

The result of the assignment is a unique MPI ID, which according to the TMF concepts (see [4]) constitutes a primary pseudonym.

1.2 Pseudonymisation

According to BDSG § 40 paragraph 2, personal data that is to be processed for scientific purposes is to be anonymised as soon as possible. According to the Landesdatenschutzgesetz Mecklenburg-Vorpommern (Data Protection Act for the German Federal State of Mecklenburg-Vorpommern [LDSG MV]) § 34 paragraph 1 , an alternative is to pseudonymise the data if there are scientific reasons for opposition to anonymisation.

Pseudonymisation (referred to in the official translation of the German Federal Data Protection Act as "aliasing") is essentially "replacing a person's name and other identifying characteristics with a label, in order to preclude identification of the data subject or to render such identification substantially difficult" (BDSG § 3, [5]). Specifically, a unique pseudo-number (pseudonym [PSN]) is assigned to the MPI ID.

The medical data is pseudonymised by the Trusted Third Party in order to:

- allow the medical data of a single person / patient / subject to be merged during a study in a clearly traceable manner,
- realise the prospective monitoring of subjects as part of follow-ups,
- allow repeated contact with the participants,
- allow quality assurance together with the institutions responsible for data collection,
- and to allow secondary-data analyses.

In the context of projects, register data retrievals and studies, the mentioned items make it necessary to allow reversal of pseudonymisation by a data custodian (the Trusted Third Party). There is a guarantee that identifiable information is only stored and processed within the Trusted Third Party. According to the TMF concepts, the PSN created constitutes a secondary pseudonym (see [4]).

Pseudonymisation is not limited to the MPI ID. Sample numbers, laboratory orders, image numbers, etc. are made reversibly indecipherable in a similar way as part of various processes within the Trusted Third Party. In order to take into account the different speciality domains, pseudonym generation is domain-specific. There is also an opportunity to re-pseudonymise existing pseudonyms (secondary pseudonymisation), e.g. in the context of data and material transfer (Transfer Office process).

1.3 Consent, authorisation and revocation

According to LDSG MV § 8 paragraphs 1-3, before data can be collected for research purposes it is necessary to "*obtain the consent of the person concerned in written or electronic form, whereby this person must be informed of 'the significance and scope of the consent, […] the type and extent of the*

processing, […] and the intended recipients' of the data."

The Trusted Third Party's other tasks include managing consents, authorisations and revocations (the informed consent [IC] forms). An IC form is modular and consists of general and project-/study-/material-specific sections.

The creation of an IC form is part of the respective project-specific ethics concept. The content is harmonised in a coordination process with the competent ethics committee and confirmed via an ethics vote or corrected under certain conditions.

Consent must include a purpose limitation. This is verified by the Trusted Third Party before the IC is recorded and before normal operations commence. Furthermore, the IC must regulate in detail study-specific aspects for which similar verification must be conducted. This includes biosamples or biomaterials, clinical data and medical device data. References are coordinated with the Data Handling unit for the respective project. These references refer to the consent modules and are used by the Data Handling unit to verify the current IC before storing medical data (MDAT). The consent must also include the following organisational aspects, which are also verified by the Trusted Third Party:

- To whom has consent been given?
- Has the patient been informed of his right of revocation?
- To whom should this revocation be addressed?
- What are the implications of a revocation?
- What form may the revocation take?

Revocation must be submitted by the patients or notifying persons to the TTP in writing. The custodian is responsible for documenting the revocation. This takes place electronically, contains a minimum of identity data and is necessary to implement the executed revocation in the case of renewed notification. The custodian is also responsible for implementing the revocation. This involves obtaining confirmation of the deletion of data in the study centres, registers, etc. involved, as well as the deletion of pseudonyms and identity data (up to the prescribed minimum for the Trusted Third Party). Finally, the revocation is logged and notification sent to the transmitter of the revocation.

In spite of the modular design of the IC form, only complete revocations are implemented. However, the technical conditions of the Trusted Third Party (e.g. in the case of separate revocation of biosample storage) allow in individual cases for revocations to be implemented on a modular basis in direct consultation with the TTP.

One consequence of the revocation is the anonymisation of data (MDAT). Although anonymisation is non-reversible and so it is not possible to cancel a revocation, renewed prospective consent is possible. Previously stored data is, however, irreversibly deleted or anonymised and biosamples are destroyed.

1.4 Working with secondary data

Another responsibility of the independent TTP is to retrieve secondary data. Potential secondary systems include civil registers, social and private health insurance companies, associations of statutory health insurance doctors and registered doctors. The Trusted Third Party supports data retrieval, comparison of secondary data and data updates.

1.5 Participating in the Use & Access process

The provision of data in the context of a Use & Access process is realised by a specially established Transfer Office at the University Medical Centre Göttingen. This Transfer Office process requires the participation of the independent Trusted Third Party (TTP).

External research projects send data and material requests to the Transfer Office. The Transfer Office determines the necessary data or material inventory using the parameters provided. Before the Transfer Office transfers the data, the Trusted Third Party checks for revocation and conducts a domain-specific secondary pseudonymisation of the data and/or material. This step is necessary in order to allow re-assignment to the person in the event of relevant incidental findings.

2 Technical systems

The technical systems (functions) of the Trusted Third Party support the data custodian in the legally compliant implementation of the processes and workflows.

2.1 Master person index (MPI)

The master person index (MPI) provides the technical functionality for the unique identification of persons described in Chapter 1.1. It is a modular software system at the Trusted Third Party that assigns a unique cross-system index (identifier) to all persons in a source domain. However, before this unique assignment can take place, the personal data is subject to a duplicate search. To do this, a configurable subset of all existing personally identifiable information is used algorithmically. At least the surname, forename, date of birth and sex are normally incorporated.

A duplicate search is necessary if the subsystems manage personal data and the associated internal local identifiers. A unique assignment of medical data to a person from several of these autonomous subsystems based on the local identifier is not possible, as the identifier is only unique within its system. The algorithm used by the MPI for detecting duplicates allows correct assignment to a unique cross-system identifier (MPI ID) even in the case of demographic information on a person that is incomplete or incorrect (to a certain extent).

The MPI extends the concept of the TMF PID generator described in [3]. The difference lies in the additional storage of domain-specific local identifiers and in the specific algorithms and weightings for the probabilistic functions used. The storage of this assignment including the local identifiers is also part of the process of re-identifying a person.

The master person index is designed as a web service. The necessary data storage is achieved using a separate MySQL database at the Trusted Third Party.

2.2 Pseudonymisation service

The Trusted Third Party's pseudonymisation service is responsible for generating and assigning pseudonyms (secondary pseudonyms in accordance with [4]) to any domain-specific character string. The service is also conceived as a web service and can be used with other services or independently. The alphabet and length of the pseudonyms, as well as the type of check digit algorithm for detecting input errors, are configurable. The individual pseudonyms are listed in a reference list within a MySQL database that is separate to but locally associated with the pseudonymisation service. De-pseudonymisation and project-specific secondary pseudonymisation (tertiary or higher level pseudonyms) are also possible.

The TMF PSN generates pseudonyms based on the identifier assigned to the person by entering a common key known as a "shared secret." The disadvantage of this approach is that all the pseudonyms must be regarded as having been compromised if the key is lost.

The PSN solution used at the TTP allows pseudonyms to be generated for any input values by use of generic algorithms. It is not necessary to enter a key value. The hypothetical knowledge of a single assignment does not imply any knowledge of algorithms or further assignments, as the generation and assignment is arbitrary and no shared secret is used.

2.3 Consent Manager

The Consent Manager is responsible for the technical management and central storage of the study participants' informed consent. Persons should be informed and enrolled by medically trained personnel (doctors) on site so that the enrolled persons have the opportunity to ask questions and receive immediate expert answers, thus providing informed consent in both ethical and legal senses. The Consent Manager indirectly supports the capture of persons' consent in electronic form through involvement in the generation of consent and authorisation forms.

A central interface is provided, via which it is possible to electronically search for study-specific consents and the associated modules (policies). Access is limited to pre-registered systems. In the context of any release of data by a transfer office, the Consent Manager allows automated daily checks on the necessary IC modules (those that correspond to the specific data requested).

The Consent Manager is also designed as a web service. The digital representations of the paper-based IC forms that remain in the enrolling study centres are assigned to the respective person at the Trusted Third Party using the MPI ID. The design of the respective IC forms is fundamentally modular. Each individual module is assigned to a technical policy. The Consent Manager also supports the version control of IC forms if they are modified or extended during the course of a research project.

Consent can only be completely revoked; revocation of individual modules is only supported in exceptional cases (see Chapter 1.3). Although the irreversible nature of anonymisation means that it is not possible to cancel a revocation, renewed prospective consent can be given.

2.4 Architecture

The architecture of the realised Trusted Third Party functionality primarily focuses on realising workflows. There are no plans for direct external access to the Trusted Third Party's individual services (MPI, CM, PSN) and such access is prevented by technical means including firewall rules and downstream, module-specific authorisation for communication partners.

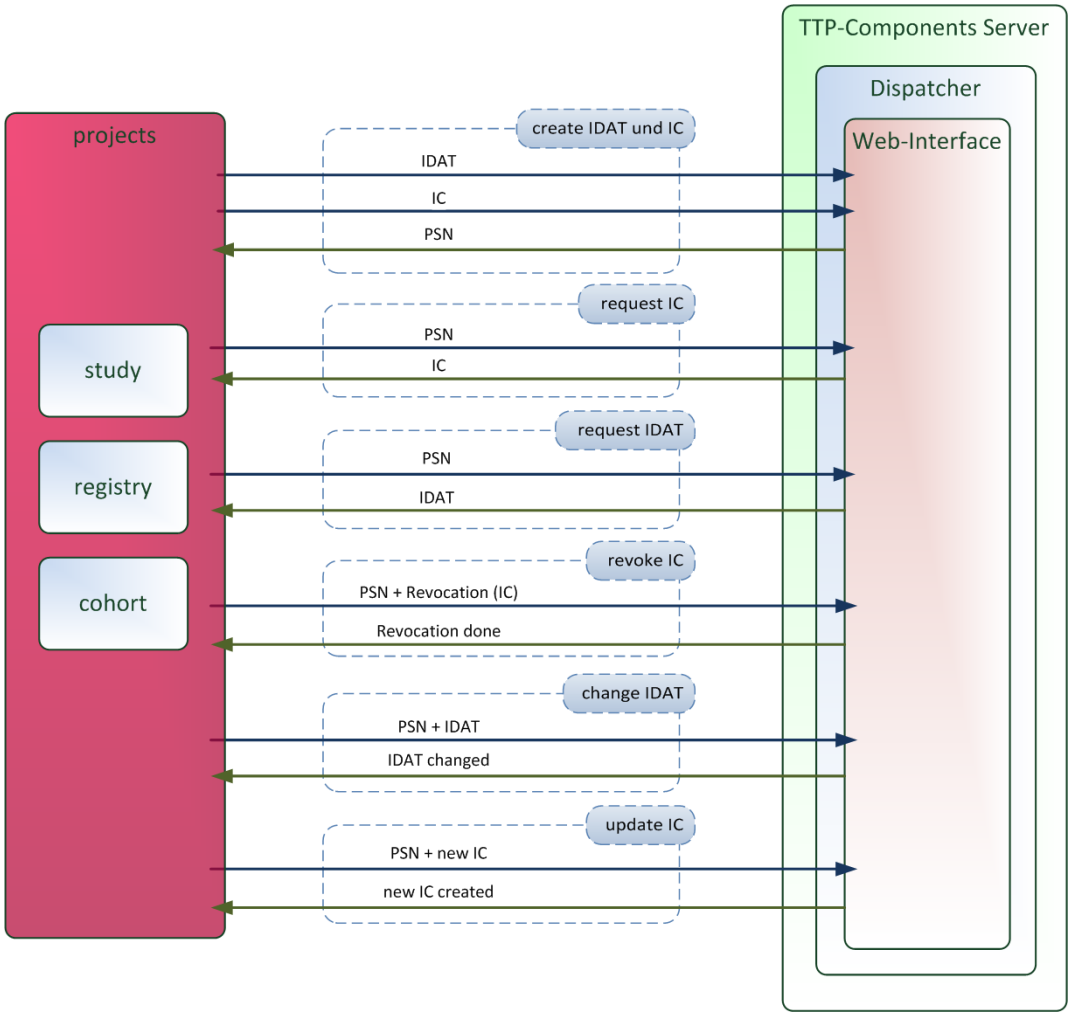


Figure 10: Possible usage scenarios for the TTP interface

Externally usable Trusted Third Party functionalities are made available by the dispatcher via a REST interface. The data is transferred in JSON format. The basic design and use concept for this interface is heavily based on the Mainzelliste [6]. Figure 10 illustrates the potential options for use of the service offered.

In addition to the individual functional modules, the central aspect of the architecture is the Workflow Manager as part of the dispatcher. The Workflow Manager manages workflow adapters, each of which contain the logic required to realise study-specific workflows. The basic functions used to implement these workflows are the TTP's MPI, CM and PSN. Every project, every study and every register, etc. has independent adapters. The Workflow Manager coordinates the respective queries sent to the underlying services via corresponding clients and sends each answer back to the system that sent the query.

The individual services are provided as a web service. Data transfer from both client and server is encrypted via HTTPS. All the necessary data is stored in separate databases, to which only Trusted Third Party employees have access. Figure 11 illustrates the relationships described.

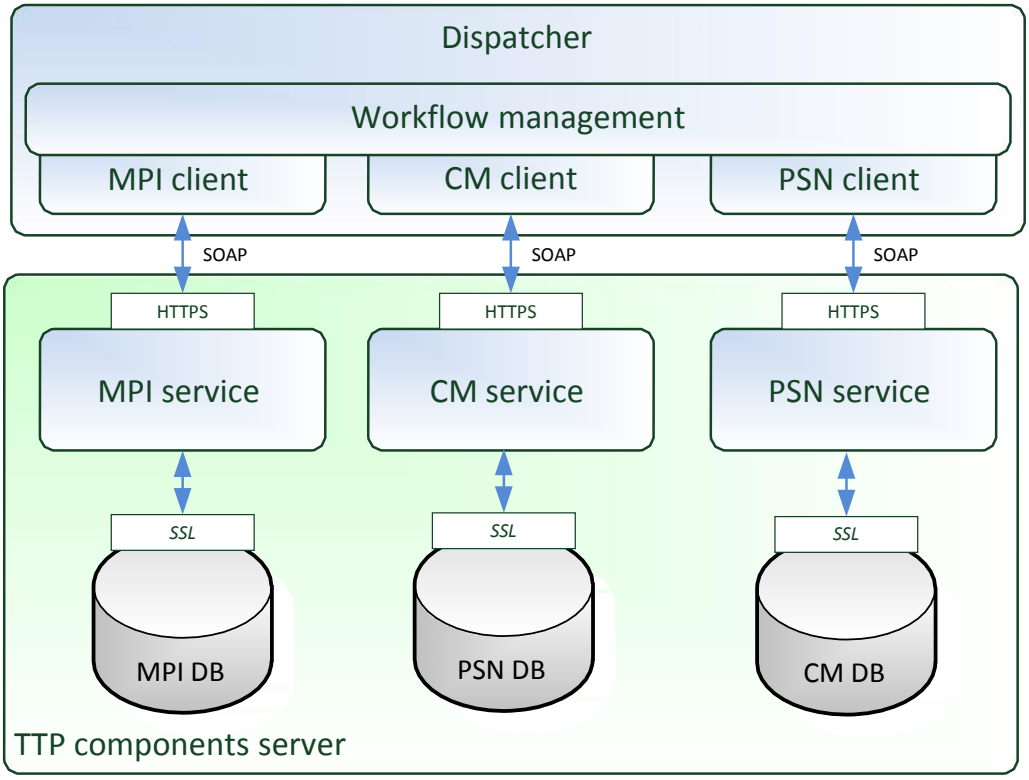


Figure 11: Technical architecture of the Trusted Third Party

3 Protection requirements

3.1 Legal basis

The processing of personal data requires compliance with a number of legal conditions. The purpose of an independent Trusted Third Party (TTP) is to ensure compliance with the relevant requirements of data protection law while allowing the flexibility necessary for research purposes and non-implementation of anonymisation. The following section describes the main provisions in order to present in detail the relevant aspects of data protection law.

Basic Law for the Federal Republic of Germany (GG)

The right to informational self-determination is derived from the general right to free development of personality and was recognised as a basic right in a census decision by the German Federal Constitutional Court. The decision stipulates that: *"Free development of personality implies the protection of the individual from unlimited collection, storage, use and disclosure of his personal data under the modern conditions of data processing. This protection is therefore covered by the fundamental right of GG Article 2 paragraph 1 in conjunction with GG Art. 1 paragraph 1. To this extent, the fundamental right guarantees the individual's general authority to decide for himself on the disclosure and use of his personal data"*[12] *[7].*

Protection of secrets according to § 203 of the German Criminal Code (StGB)

Projects with a medical context must comply with the duty of medical confidentiality, which is regulated in StGB § 203.

This duty of confidentiality applies not only to doctors, pharmacists and similar medical professions, but also to those who support these persons. These persons include nursing staff and medical technologists. [7]

Data protection law

Whether an institution is subject to the German Federal Data Protection Act (BDSG) or the corresponding federal state law is decided on the basis of the respective legal form ([8], p. 145).

Non-governmental organisations, federal authorities and private companies are subject to the collection and processing of personal data according to the BDSG. [8]:

- The principles of data reduction and data economy apply (BDSG § 3a).
- Admissibility exists through appropriate legislation or the consent of the person concerned (BDSG § 4).
- Data protection is implemented using appropriate technical and organisational measures.

For the Trusted Third Party, as the data-processing unit at University Medicine Greifswald (a public body in the German federal state of Mecklenburg-Vorpommern) this specifically means that the collection, processing and use of medical research data is subject to the LDSG MV.

As the TTP is involved in several national projects and studies, clarification is required on which German federal state law is applicable in the case of cooperation between parties in different German federal states. This is based on the respective authority. **If the data is processed in the TTP on behalf of a project or a study, it receives instructions, does not make any decisions and is in this case subject to the data protection law of the German federal state in which the client is based. If, however, functions are transferred, data processing is the sole and exclusive responsibility of the TTP and is therefore subject to the LDSG MV.**In this case, the main point of contact for data protection affairs is the Landesbeauftragter für den Datenschutz und die Informationsfreiheit Mecklenburg-Vorpommern (State Commissioner for Data Protection and Freedom of Information Mecklenburg-Vorpommern [LfD MV]). Furthermore, project partners have the opportunity to impose additional conditions on the agreed transfer of functions, e.g. compliance with data protection conditions specific to a particular German federal state.

[12] Decision of the German Federal Constitutional Court (BVerfGE) 65, 1 – Census, 1983

The LDSG MV regulates in detail all steps for processing personal data (LDSG MV § 7-§ 23). For example it stipulates that:

- the personal data is to be anonymised as soon as possible (§ 5). Pseudonymisation is also a permissible alternative.
- the consent of the person concerned is to be obtained in written or electronic form, whereby the person concerned is to be comprehensively informed of 'the significance and scope of the consent, [...] the type and extent of the processing, [...] and the intended recipients' of the data" (LDSG MV § 8 paragraphs 1-3).
- the use of data is only permissible for the purpose for which it was collected (§ 10 paragraph 2). Any other use requires the consent of the person concerned (§ 10 paragraph 3).
- incorrect data must be corrected or deleted (§ 13 paragraphs 1 and 2). It is also possible to lock the data (§ 10 paragraph 3).
- in the case of data transfer to other parties, the transferring party is responsible (admissibility, data protection, IT security) (§ 14-§ 16).
- specific measures are required to maintain data security (§ 21).

At the same time the LDSG MV regulates the rights of the person concerned with regard to information, locking and revocation (§ 24 and § 25) and stipulates special requirements for scientific research (§ 34). The technical, staff, spacial and organisational measures necessary for implementation of the required conditions are stipulated in a data protection concept and harmonised with the LfD MV.

State hospital law

State hospital law stipulates additional requirements for patient data protection[13].The Landeskrankenhausgesetz Mecklenburg-Vorpommern (Hospital Act for the German Federal State of Mecklenburg Vorpommern [LKHG MV]) § 38, for example, regulates data processing for research purposes. Paragraph 1 requires the consent of the patient for the use of patient data for research purposes. Paragraph 2 defines exceptions for which the consent of the patient is not required, e.g. if interests worthy of protection are not compromised because of the type of data or because of the evident use of the data, or if the public interest in conducting the research project outweighs the interests of the patient worthy of protection. Furthermore, paragraph 4 requires that data is pseudonymised (clause 1) or anonymised (clause 2) as soon as possible. [7]

3.2 Storing personal data

The medical context means that the legal conditions require increased sensitivity when handling personal data. The aforementioned data protection laws regulate all phases for the processing of medical research data. At the same time, they include possible consequences associated with violating the rules and restrictions.

[13] The distribution of paragraphs differs depending on the law of the respective German federal state. LKHG MV § 32-§ 39, for example, regulate patient data protection.

In order to realise epidemiological and research-related analyses on the basis of collected research data, it must be possible to assign a person's medical and identifiable information both uniquely and as accurately as possible (see Section 1.1, Chapter 1.1), while also complying with data protection requirements.

Identifiable information (IDAT) comprises the surname, forename, maiden name (if different), sex, date of birth, place of birth, address and contact details such as telephone, fax and email (see Table 2). The IDAT dataset can be expanded as part of a study after consultation with the Trusted Third Party. IDATs are separated from medical data (MDAT) as soon as possible; the Trusted Third Party never knows the medical data (MDAT) of the person concerned. However, the specific requirements of the IDAT are dependent on the respective design of the study. The person's place and date of birth are used for precise assignment of the pseudonymised data and as part of scientific evaluations. The address and contact information are generally used to contact the patient at a later date (if consent has been given).

In individual cases it is also possible to store additional personal data, such as case numbers, laboratory order numbers and image numbers. These are subject to the same stipulated regulations. The Trusted Third Party assumes responsibility for storing the required informed consent forms, which comprise consents, authorisations and revocations.

The duration of the data storage is defined by the respective study or register in the informed consent form and ethics vote. The legal provisions must be taken into account when stipulating the duration.

IDAT	Purpose
Surname, forename, sex, date of birth, place of birth, address, email, telephone, fax	Identification of the person within the Trusted Third Party for unique assignment of his MDAT Register data retrieval, secondary-data comparisons and merging with secondary data, follow-ups (e.g. vital status), assistance in re-establishing contact
Case, laboratory order and image numbers	Unique assignment of materials to the person
Electronically evaluable consent of the participating person	Basic requirement to store and retrieve the IDAT within the Trusted Third Party
Information on the project/study/register collecting the data	Quality assurance Re-establishing contact
Collector's identifiable information: surname, forename, title, department	Quality assurance Re-establishing contact

Table 2: Overview of the personal data stored at the TTP

3.3 Determining protection requirements

The personal data to be stored is individual information regarding the personal circumstances of a specific or identifiable natural person. This means that the protection requirements for the confidentiality of the data to be stored are high. The following illustrates the protection requirements for the Trusted Third Party's systems on the basis of basic requirements for IT security and in terms of general data security measures (LDSG MV § 21). Protection requirements are classified according to three categories (normal, high and very high) (see [9]).

Baseline value	Protection requirement	Explanation	Reason
Confidentiality	High	The data may only be seen and modified by authorised users. This applies both to access to stored data and to the data transfer	Within the Trusted Third Party, persons' IDAT is processed in accordance with the applicable German federal state data protection act (LDSG; see LDSG MV §§ 24, 25 and 34; accordingly, the corresponding paragraphs of the respective LDSG apply). Assignments between the MPI ID and PSN are stored. This makes it possible to trace the identity of participants of a project / study. Compromised confidentiality through unauthorised access has a negative effect on public trust. This can lead to reduced participation in studies and can therefore also influence the quality of studies.
Availability	Normal	Prevention of system failures; access to data within an agreed time frame must be ensured (in accordance with GEP (Good Epidemiological Practice), GCP (Good Clinical Practice) and IT/ information security (IS) of the LDSG).	Failure on the part of the TTP can compromise the functional capacity of the study centres or registers.
Integrity	High	Personal data must remain undamaged, complete and up-to-date during processing.	Loss of data integrity as a result of incorrect or incomplete assignment results in financial losses for the project and renders the data useless.

Authenticity	High	It must be possible to verify the authenticity and credibility of a person or service	It must be possible to assign personal data to its origin at any time. If it is not possible to guarantee the correctness of the data, its integrity and the person's right to free determination of personality are compromised. Furthermore, incorrect assignment (e.g. in the case of incidental findings) can result in incorrect treatment and cause considerable personal and mental distress.
Traceability	High	It must be possible at any time to determine who processed what personal data at what time in what manner.	Compromised tracing security can compromise the right of informational self-determination and thus damage the public reputation of and public trust in the TTP. Complete traceability can help to clarify liability issues by way of logs.
Transparency	High	It must be possible to trace processes completely and temporally as far as is reasonable.	Only complete documentation makes it possible to realise the legal requirements for provision of information. In this case, a loss of transparency would constitute a violation of the law and have negative consequences. In the event of an error, complete transparency can allow the error to be identified and issues of liability to be clarified. A loss of transparency leads to a limited documentation of data processing. Thus, the accountability of data processes is influenced.

Table 3: Determining the protection requirement of Trusted Third Party systems (according to [7])

4 Technical and organisational measures

4.1 ICM-VC institutional data protection concept

The organisation, spacial separation, staffing and technical equipment of the Institute for Community Medicine at University Medicine Greifswald are described in the "Rahmenkonzept Datenschutz und IT-Sicherheit für das Institut für Community Medicine der Universitätsmedizin

Greifswald" (Framework concept for data protection and IT security for the Institute for Community Medicine of the University Medicine Greifswald) [10].

The data protection and IT security regulations defined in the framework concept have been coordinated with the LfD MV, describe the principles of working with data at the Institute for Community Medicine, Section Epidemiology of Health Care and Community Care (ICM-VC) and are thus considered the standard basis for project-specific aspects of the Independent Trusted Third Party (TTP) operating at the ICM-VC. These are clarified and expanded upon in the following Chapter A.

4.2 Network protection

The Trusted Third Party and the services involved operate in an isolated network area (the TTP zone) that is separated from the University Medicine Greifswald network.

In order to further improve security, a *Drei-Zonen-Konzept* (*Three Zone Concept*)[14] in accordance with the BSI's catalogue of basic protection and measures entitled "*M 5.117 Integration eines Datenbank-Servers in ein Sicherheitsgateway*" *(M 5.117 Integrating a database server into a security gateway)*[15] has been implemented on the initiative of the Institute for Community Medicine at University Medicine Greifswald. Web services and related databases are operated in separate zones.

It is possible to access the Demilitarised Zone (TTP DMZ) from the internet ("External" Zone). The connection to the Transfer Zone (TTP TZ) is based on the DMZ. Access from the Militarised Zone (TTP MZ) to the Transfer Zone is also permitted. "Penetration" via the TTP TZ is not possible: connections can only be made into the TTP TZ; they cannot be actively established out of it. The communication flow regarding the direction of connection establishment is limited by the University Medicine Greifswald firewalls. Closed, project-specific zones exist within the entire network. Connections can only be established between the individual sub-networks in the context of the respective project or use.

Figure 12 illustrates the relationships described. Access to TTP services is only possible via encrypted connection after successful authentication. Communication between projects and TTP services is realised via zone-specific proxy servers and is only available to registered IP addresses and ports.

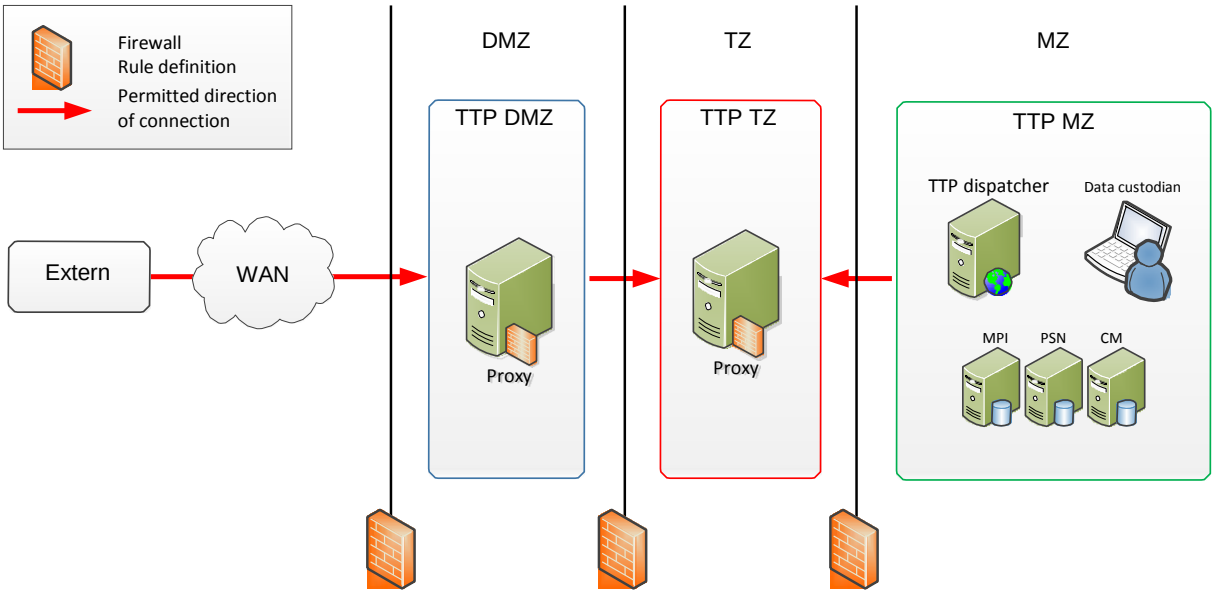


Figure 12: Trusted Third Party Zone Concept

[14] See the attachment entitled "Konzept für den sicheren internen und externen Zugriff auf Forschungsdienste" (Concept for secure internal and external access to research services)

[15] https://www.bsi.bund.de/DE/Themen/ITGrundschutz/ITGrundschutzKataloge/Inhalt/_content/m/m05/m05117.html

4.3 Audittrail

Within the Trusted Third Party, all access to systems and data is documented via an audit trail process. All subsystems log who accesses the system, when the access takes place and which particular data is accessed. This is valid for both external and internal access (e.g. access by the TTP's technical staff). This means that all changes made to data can be traced and that the integrity of the data is guaranteed at all times. Various versions of data status can also be restored in the event of an error.

4.4 Data transfer

HTTP is the standard transfer protocol used to transfer data. External systems must use the encrypted HTTPS version with 256-bit encryption (or higher) for data transfer. In order to guarantee a correspondingly high level of security, HTTPS with TLS Version 1.0 encryption or higher is used. Browsers that only support TLS 1.0 should also use a client certificate. Browsers that use TLS 1.2 or higher are regarded as secure even without a client certificate. RSA with key sizes of 2048 Bit, AES-256 and SHA-256 are used as algorithms. These correspond to the recommendations of the Bundesnetzagentur[16], the network agency for the German federal government, and are considered secure until the end of 2018.

Communication between the Trusted Third Party and external systems is also protected by restricting IP addresses. Only authorized IP addresses and ports are permitted to connect to TTP services (see the Zone Concept [Chapter 4.2]).

4.5 Data security

Data security is realised via the implementation of a rights concept on the basis of access levels. All Trusted Third Party servers are encrypted (AES-512), i.e. encryption is performed independently by the local operating system. All communication via the network is encrypted either automatically or upon request. The data custodian has the necessary password. A copy of this password can be found in the bank deposit box of the Trusted Third Party. Only the Head of the Trusted Third Party and the data custodian have access to this box.

Every week, authorised administrators perform a complete backup of the servers to protect against natural hazards. However, the encryption means that these administrators cannot view the data at any time. The backups are stored on a separate tape drive in the same network segment and is then placed in the Trusted Third Party's bank deposit box in accordance with the institution's security concept [10]. There is also a daily backup of the systems on a separate tape drive for the purpose of disaster recovery.

16

https://www.bundesnetzagentur.de/SharedDocs/Downloads/DE/Sachgebiete/QES/Veroeffentlichungen/Algorithmen/2012Algorithmenkatalog.pdf

4.6 Failure protection

In order to guarantee proper operation by external partners, the services provided by the Independent Trusted Third Party (TTP) must be permanently and reliably accessible online. The following illustrates measures that increase system availability and minimise downtime. Possible causes for system non-availability are listed in the column marked "Cause."

Cause	Measure
Faulty system configuration (software)	Testing for configuration changes on the test system Logging of the changes made using suitable software Automated monitoring of services using suitable software (e.g. Nagios or Xymon) Changes and adjustments are conducted only by employees who receive continuous training
Faulty system configuration (hardware)	Before production operation system and load tests are performed to ensure that the server is functioning correctly. The same tests are conducted after components are replaced.
Individual component failure (hardware)	Operation of the system on two parallel servers in a high-availability cluster
Server system failure (hardware)	Operation of the system on two parallel servers in a high-availability cluster
Power supply failure (infrastructure)	Connection of the server to an uninterruptible power supply (UPS) for bridging short-term power supply failures The fact that the system is not absolutely necessary for medical care means that there is no emergency generator to supply power in the event of longer-term power supply failures (subsequently documented in the system when power is available again).
Attack on server	Use of restrictively configured firewalls Comprehensive monitoring of authentication and unusual data requests

Table 4: Failure protection measures

In addition to the measures described above, a failure assessment process starts if non-availability exceeds six hours in the case of both unplanned and planned (e.g. an update with a system re-start) limited availability. The duration of the limited availability, the reason for the limited availability and the effect of the limited availability on the usability of the system are then recorded in basic documentation. Technical and organisational measures are then developed in order to prevent limited availability in the future or to identify them earlier and possibly remove them automatically.

4.7 Spacial separation

The Trusted Third Party is based in a separate building located near the Epidemiology of Health Care and Community Section. This ensures without restriction that the measures and processes stipulated in the Rahmenkonzept Datenschutz und IT-Sicherheit für das Institut für Community Medicine der Universitätsmedizin Greifswald (Framework concept for data protection and IT security for the Institute for Community Medicine, University Medicine Greifswald) are guaranteed during TTP operation.

The TTP also has a separate entrance, a separate closed circuit and its own alarm system. The locked premises may only be accessed if the appointed data custodian is present.

4.8 Personnel measures

The data custodian is responsible for the organising aspects of the Trusted Third Party, is not assigned to any particular project and is the paramount authority in relation to project partners or the Institute for Community Medicine (and the superior University Medicine Greifswald) in which s/he is based. According to the law (LDSG MV), the data custodian is a defined group of TTP employees who, in their function as data custodians, are subordinate only to the Head of the Trusted Third Party. The Head of the Trusted Party has the paramount authority in his role as data custodian.

There is currently no legal regulation on establishing a data custodian, but it has been recommended since 2000. [11] [12]

B2 Data handling unit

5 Data handling processes

The Data Handling unit works with SOPs internally, which address the creation of users, a controlled software update process and further items necessary for secure operation of the application. Up-to-date versions of the following SOPs are currently used:

Table 5: Overview of the SOPs in use

No.	Internal designation	Description
1.	MI-sT01_SOP_Struktur Erstellung Änderung	Description of the SOP structure and the procedure for amendments (Master SOP)
2.	MI-sT02_ Systemverwaltung	Instructions for the management and operation of the IT infrastructure used in the MI
3.	MI-sT03_SOP_eCRF-Erstellung	Procedure for creating eCRFs and for conducting user and system tests
4.	MI-sT04_SOP_PID-Erstellung und Verwaltung	Creating PIDs and managing pseudonymisation of research data
5.	MI-sT05_SOP_ Benutzerverwaltung	Creating, amending, disabling and deleting user accesses
6.	MI-sT06_SOP_ Export und Audit-Trail	Authorising, documenting and storing exports
7.	MI-sT07_SOP_ Datenbank sperren und wiederöffnen	Locking and, if applicable, reopening a database after completion of the study
8.	DM08_SOP_Systemversionierung	Instructions for system version control
9.	MI-sT09_SOP_ Langzeitarchivierung	Long-term archiving of data after completion of the study Backing up data after completion of a project Archiving data and disabling clients
10.	MI-sT10_SOP_ IT-Fehlermanagement	Description of processes to avoid errors and of how to respond to errors in the event of occurrence
11.	MI-sT11_SOP_Validierungen im laufenden Betrieb	Description of processes to avoid errors and of how to respond to errors in the event of occurrence

12.	MI-sT12_SOP_ Dokumentenmanagement	List of documents required for data management, documentation of decisions in software systems for projects
13.	WI01_Systemsicherung und Notfallmanagement	Back-up procedure, emergency and damage management
14.	MI-sT13_Schulungen	Training on new SOP content and user training
15.	MI-st14_SOP_Löschen von Patientendaten	Deleting patient data / Revoking informed consent

With consistent standard operating procedures (SOPs), the Data Handling unit specifies required and binding rules. The requirement for the existence of SOPs is an integral part of the International Conference on Harmonisation (ICH) Harmonised Tripartite Guideline of Good Clinical Practice (GCP), Chapter 5.1. The subject of the SOPs is processes within the electronic data management of clinical studies within the Data Handling unit that require standardisation and recording in written form due to their complexity or security relevance. The SOPs take into account the requirements of data management and the collection of electronic casebooks in Chapter 5.3 of the International Conference on Harmonisation (ICH) Harmonised Tripartite Guideline for Good Clinical Practice (GCP).

6 Technical systems

6.1 secuTrial

The DH unit's core task is to provide a data capture tool that is easy to use but also complies with all data protection conditions. secuTrial is used for this purpose. With secuTrial it is possible to conduct multicentre clinical studies and post-marketing studies. secuTrial allows direct, decentralised electronic capture of study data (remote data entry) in a central database.

Operation of secuTrial is completely browser-based, which means it is not necessary to install software for either management or data capture. Authorised users can define the studio setup, manage participants, export data and enter patient data from any internet-enabled PC.

Within the application there is both a separate test area (setup) and a productive area. This means that prior to the start of the actual study or in the case of changes implemented after the go-live, any function can be tested before it is unlocked for users. The tested study setup can go live after it has successfully completed testing. The changes are subject to constant version control.

secuTrial complies with all regulatory standards (CRF, GCP) and is certified for all FDA-compliant functions, such as audit trail, a roles and rights concept and electronic signature.

secuTrial has also repeatedly undergone independent benchmarking audits and has been certified to be in complete compliance with 21 CRF Part 11 and the regulations based upon it.

Internal structure of the application

Within secuTrial there are five different tools that guarantee that specific administration tasks are fully covered by operative tasks on the part of the data capturer. This modular design is illustrated in Figure 13.

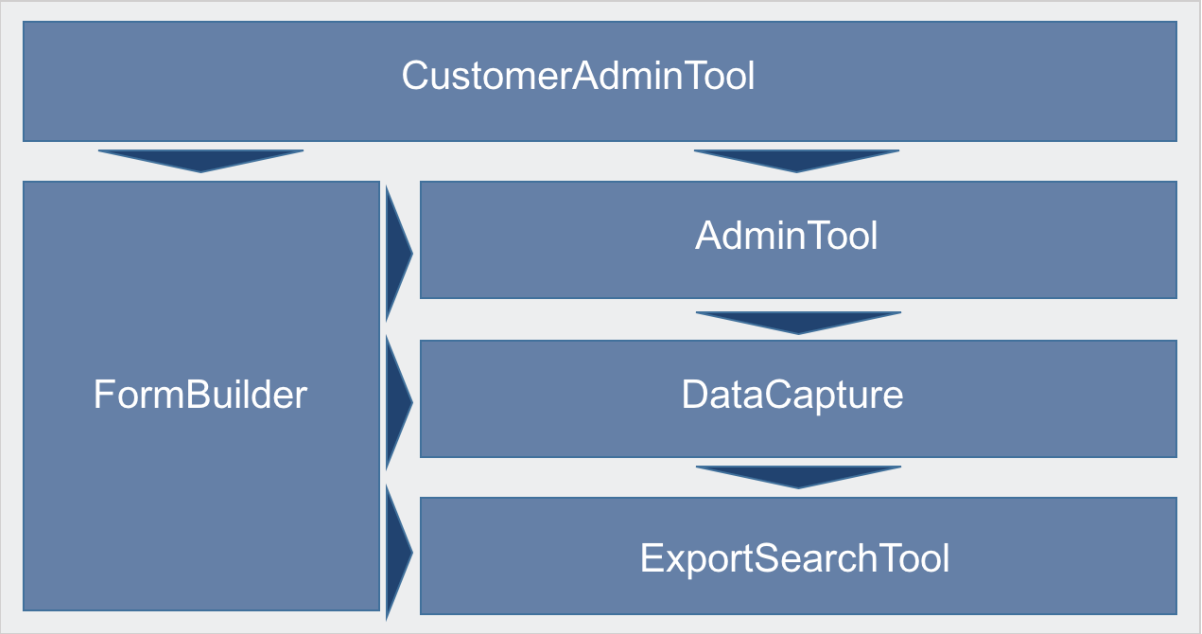


Figure 13: Modular design of secuTrial

The overall system consists of the following individual modules:

1. *CustomerAdminTool*
   Function: Managing customers and administrators, creating project schematics (DB areas), generating statistics, sending bundled messages, controlling the file system (untitled images, temporary data), archiving and deleting, DB documentation
   Access check: User in password file
2. *FormBuilder*
   Function: Creating and configuring projects, creating forms, project version control, internal changes to the database
   Access check: Participants from AdminTool management
3. *AdminTool*
   Function: Participant and patient management, roles and rights management, changes to design, customer-specific user guidance
   Access check: Administrator from CustomerAdminTool management

4. *DataCapture*
   Function: Creating patients, capturing data, data overviews (reports and statistics), managing queries, importing data
   Access check: Participants or patients (only in the case of patient self-documentation) from AdminTool management
5. *ExportSearchTool*
   Function: Searching for patients, exporting data Access check:
   Participants from AdminTool management

7 Protection requirements

In medical research, various personally identifiable and non-personally identifiable information is captured, processed and stored. Various preparations must be made in order to protect the rights of patients. The first step is to obtain the informed consent of the patient (or of the patient's legal guardian). If the patient agrees to the processing and disclosure of his data and/or the sampling and forwarding of biomaterial, the data and biomaterial may be used for research. Access to the medical records of the study participant is also regulated in the informed consent form. The medical data collected for research is processed in pseudonymised form and organisationally and technically separately from personal data. Personal data is processed by a data custodian, while the pseudonymised medical research data is processed by the Data Handling unit. The Data Handling unit in Göttingen cannot legally trace back the pseudonymised medical research data to the patient's personal data.

The technical and organisational measures described in NDSG § 7 for processing personal data are also used for processing the pseudonymised medical data. This is partly achieved by the fact that the secuTrial application is used by the Information Technology business unit at the University Medical Centre Göttingen, which is also responsible for the operation of the applications in patient care. It is also achieved by the fact that the Department of Medical Informatics at the University Medical Centre Göttingen works with SOPs that stipulate the use of the secuTrial application (see Section 5).

7.1 Processed data types

Only pseudonymised medical data is processed within the Data Handling unit. Based on the architecture, identifiable information is processed only by the data custodian. The medical data processed in the DH unit includes the results of patient interviews in the form of completed eCRFs. These eCRFs are divided into several modular datasets. They are completed based on the specific study. The datasets are essentially identical. The phenotypic database contains clinical data on the individual study participants collected as part of interviews and various self-assessment and clinical assessment scales. The individual study design determines what data is collected and what scales are used.

7.2 Applicable legal basis

Landesdatenschutzgesetz Niedersachsen (Data Protection Act for the German Federal State of Lower Saxony [NDSG])

NDSG § 25 *Processing personal data for research purposes* applies to data handling. The items it stipulates are all taken into account. This means the written consent of persons concerned is obtained. Furthermore, the data stored and transferred for the research project is only used for the purpose of scientific research. Characteristics with the help of which a connection to a certain natural person can be made are also stored separately. Data transfer is only permissible if the recipient also only uses the data for the stipulated research project. This intended use is correspondingly recorded in the informed consent form.

German Federal Data Protection Act (BDSG)

All relevant requirements of the German Federal Data Protection Act (BDSG) are taken into account within the Data Handling unit. Data that is collected within the studies is treated confidentially and only used for the purpose stated in the informed consent form. The data can be disclosed to third parties in pseudonymised or anonymised form if required. The data is only saved locally in pseudonymised form. [5] At any time, the subjects have the right to receive information about their data or to revoke their consent. If subjects revoke their consent, the data and biomaterial is anonymised or destroyed, if applicable, so that the captured data can longer be assigned to the identifiable information. All data is pseudonymised in accordance with BDSG § 3 paragraph 6a. This means that the name and other identifying characteristics of the subject are replaced with an identifier for the purpose of preventing or substantially complicating the identification of the person concerned.

German Medicinal Products Act (AMG)

"It is the purpose of the present Act to guarantee, in the interest of furnishing both human beings and animals with a proper supply of medicinal products, safety in respect of the trade in medicinal products, ensuring in particular the quality, efficacy and safety of medicinal products in accordance with the following provisions." [13] The applicability of this law derives from the specific studies conducted and only comes into effect if medicinal products are investigated as part of these studies.

Gesetz über Medizinprodukte (German Medical Devices Act [MPG])

"It is the purpose of the present act to regulate the trade in medical devices and thereby guarantee the safety, suitability and performance of medical devices as well as the health and required protection of patients, users and third parties." [14] The applicability of this law derives from the specific studies conducted and only comes into effect if medical devices are investigated as part of these studies.

Verordnung über die Anwendung der guten klinischen Praxis (German Good Clinical Practice Ordinance [GCP-V])

According to GCP-V, planning, conduct, documentation and reporting must be conducted as part of the studies. [15] This ensures that the rights, safety and well-being of study participants are protected in accordance with the Declaration of Helsinki.

8 Technical and organisational measures

8.1 IT infrastructure used

The components required to operate the secuTrial study database (database system and application software) are installed on virtual machines on servers in a restricted-access server room belonging to the Information Technology business unit at the University Medical Centre Göttingen (UMG) under the management of Prof. Dr. Ulrich Sax. The phenotypic database is located in the WissLAN[17] network segment.

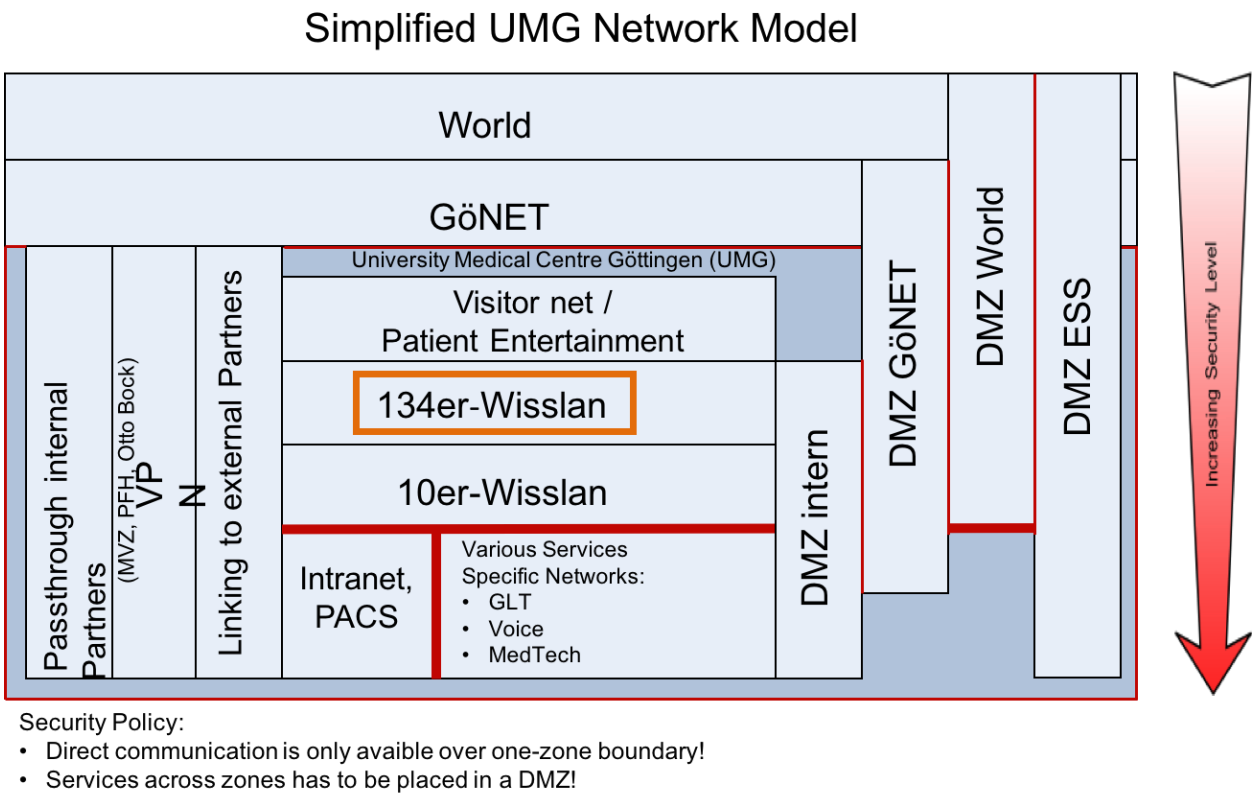


Figure 14: Illustration of the current network infrastructure status at the UMG. The separation of the internal segment for patient care (in this case, intranet and PACS) from the other network segments has proven itself and is consistently applied for reasons of security. The infrastructure of the Department of Medical Informatics is located in the *134er-WissLAN* network segment.

[17] WissLAN is a network segment within the UMG and is dedicated to scientific purposes. It specifically differs from PatLAN, which is exclusively used for patient care.

Within WissLAN there is a dedicated research infrastructure for the Department of Medical Informatics. This is separated from the public part of WissLAN by a two-stage firewall, thereby fulfilling basic IT protection requirements in accordance with the German Federal Office for Information Security (BSI) Recommendation [16]. BSI Measure *M 4.223 Integration von Proxy-Servern in das Sicherheitsgateway (Integration of proxy servers into the security gateway)* [17] has been implemented. The dedicated research infrastructure implements the *reverse proxy* construct of BSI measure *M 4.223*. Communication within the protected network is not encrypted. The external firewall is configured as a packet filter. The reverse proxy encrypts all communication with all public networks. The architecture is illustrated in Figure 15.

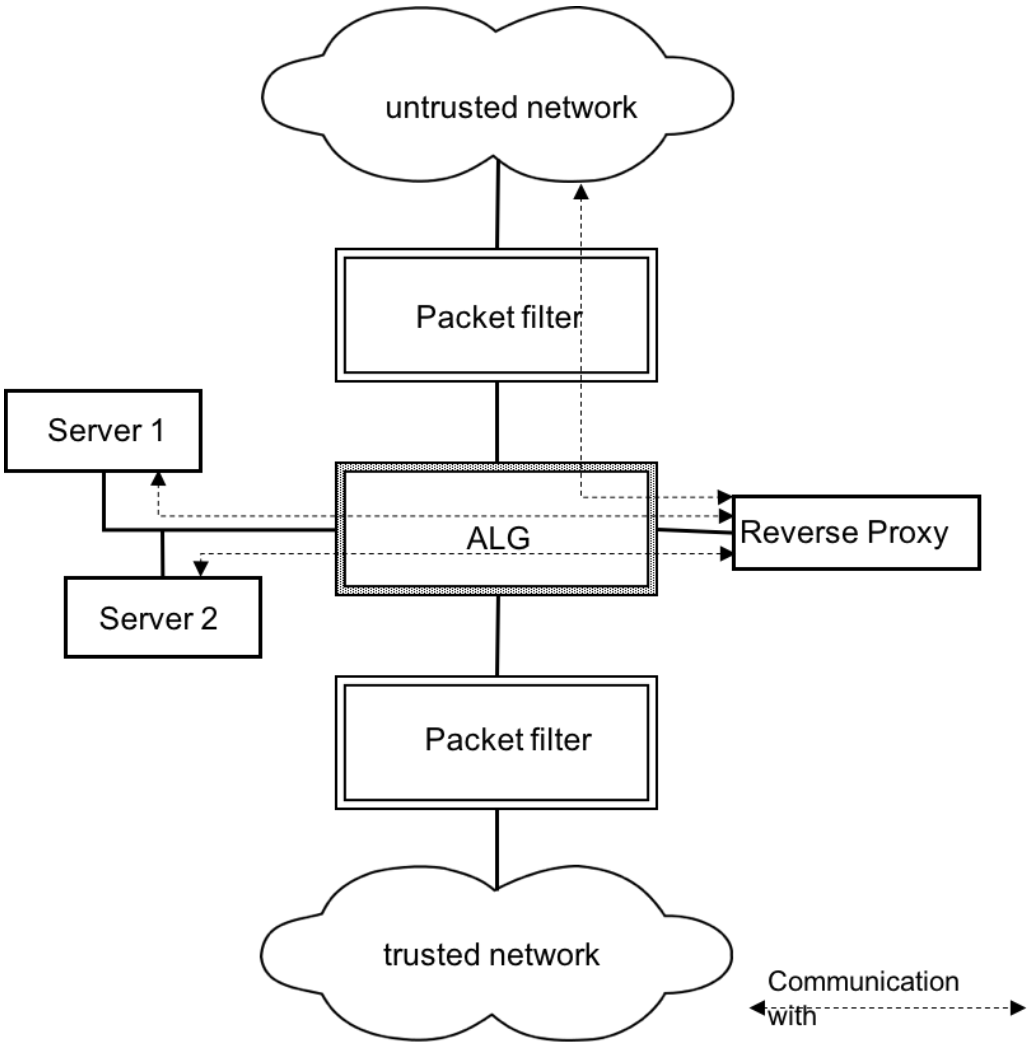


Figure 15: Schematic illustration of the dedicated research infrastructure The architecture is based on BSI Measure *M 4.223* [17].

8.2 Server virtualisation

Server virtualisation technologies are used within the dedicated research network. These comply with the recommendations of the Arbeitskreis Technik (Technology Working Party) of the Konferenz der Datenschutzbeauftragten des Bundes und der Länder (German Conference of Commissioners for Data Protection of the Federal Government and the Federal States [DSK]) [18].

8.3 Process description in accordance with NDSG § 8

The description of the process used in the Data Handling unit can be found in the attachments. This process is coordinated with the Commissioner for Data Protection of the University Medical Centre Göttingen. This document serves to achieve transparency and access to information for the person concerned, as well as traceability. It is therefore necessary to record in this document which personal data is processed using which automated procedure in which manner and which data protection measures were taken.

C Annex

1 Technical illustrations

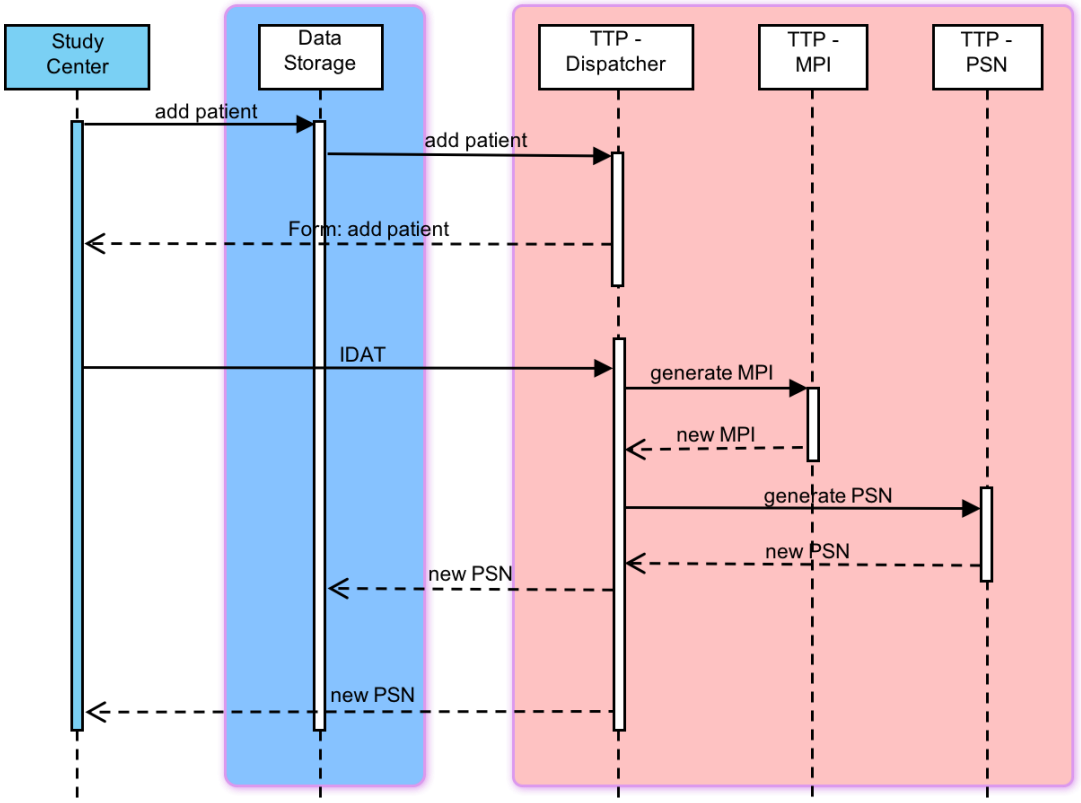


Figure 16: Creating a new study participant (technical view)

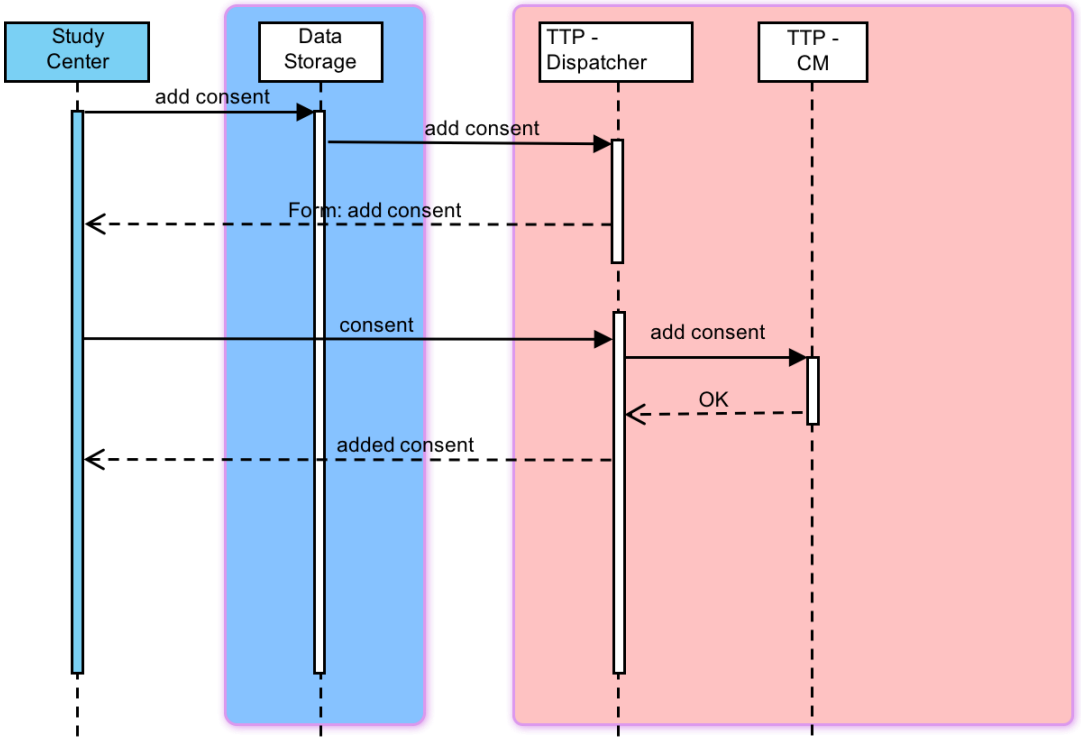


Figure 17: Creating an IC form (technical view)

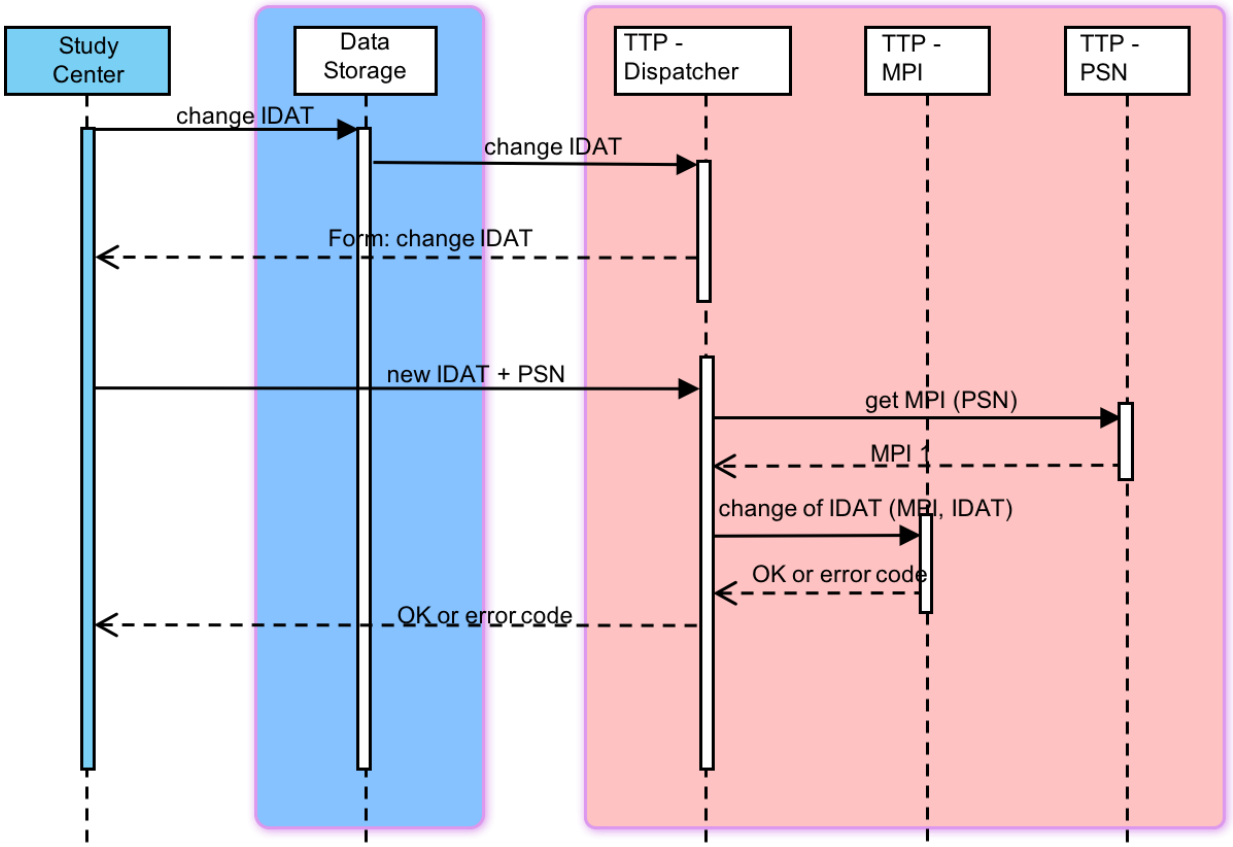


Figure 18: Amending IDAT (technical view)

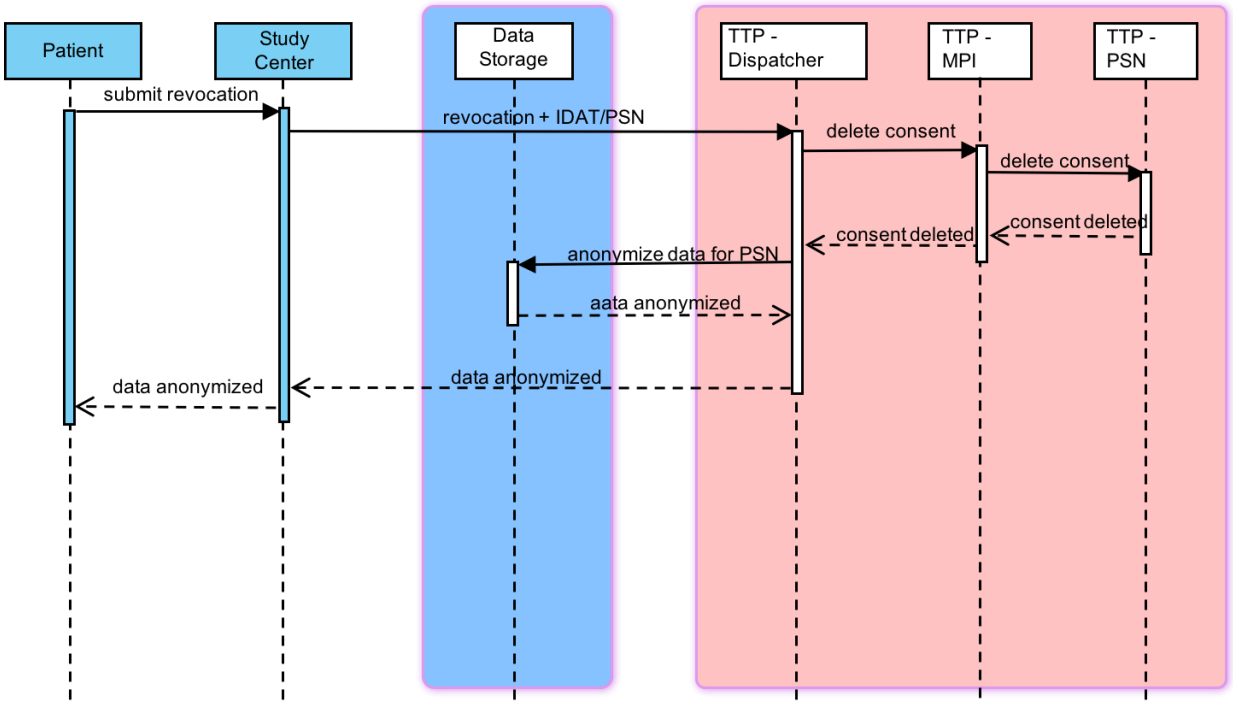


Figure 19: Revocation (technical view)

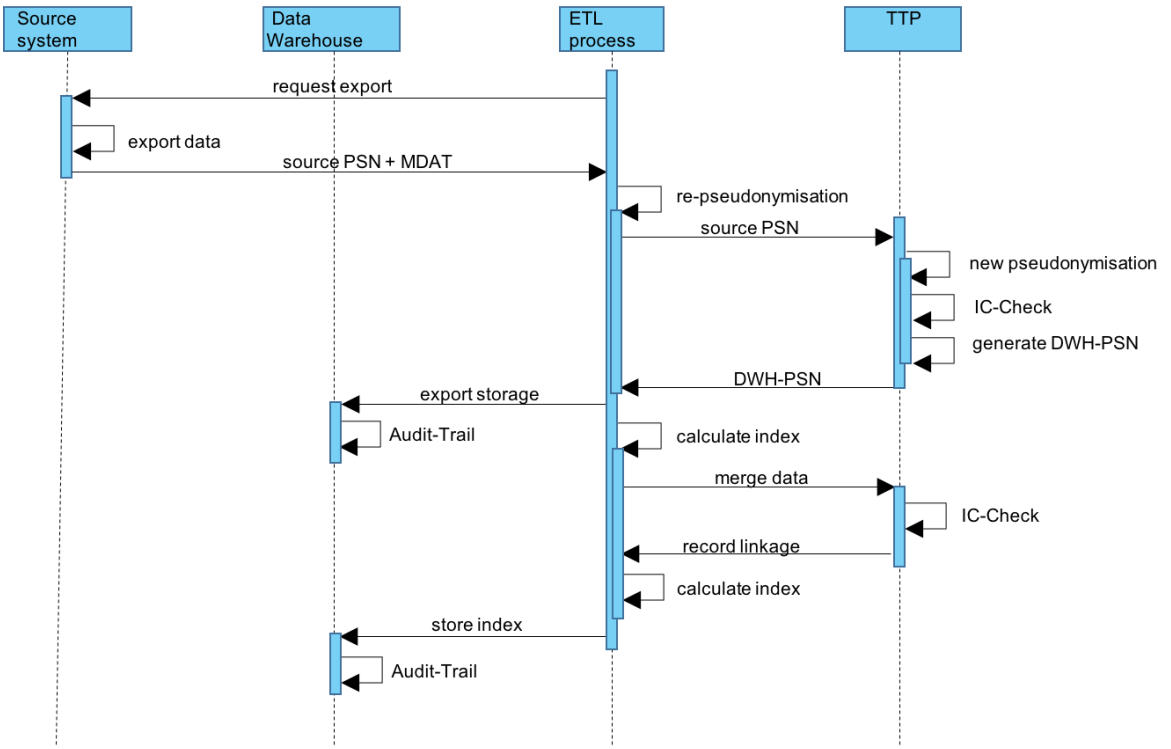


Figure 20: Technical process for exporting and re-pseudonymising medical data from a source system (e.g. secuTrial) with the connected ETL process and interaction with the TTP

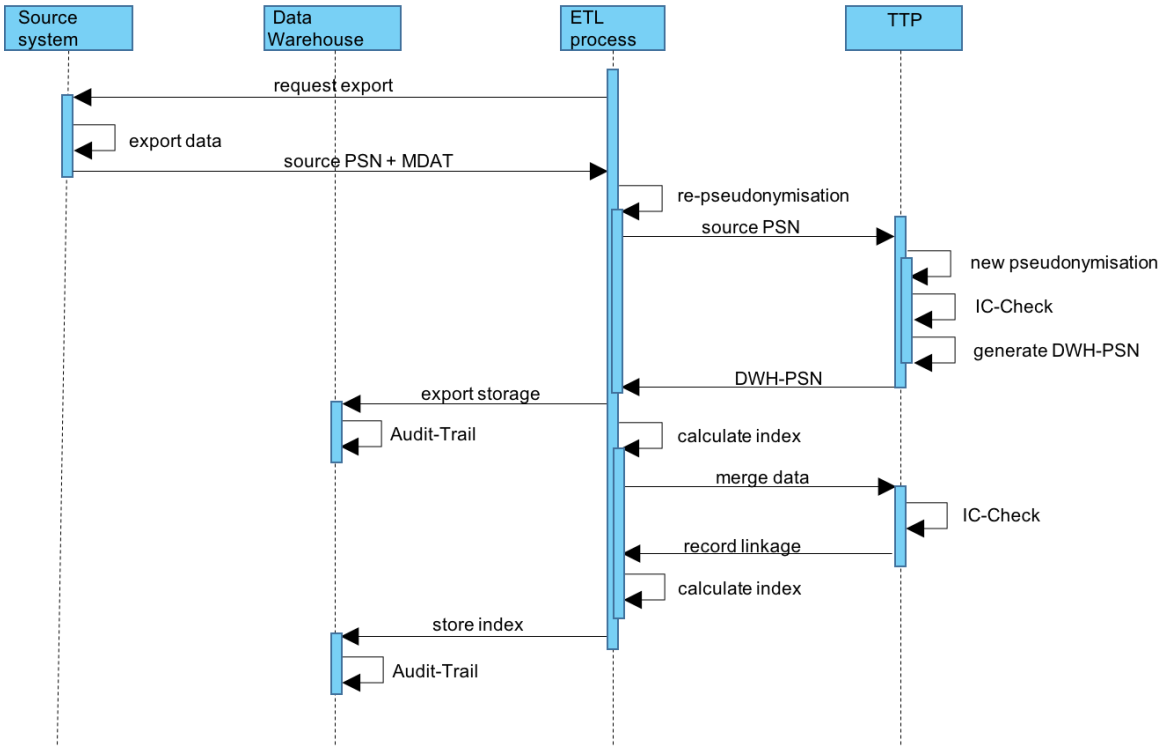


Figure 21: Organisational process for registering and using quality management quantitative indicators

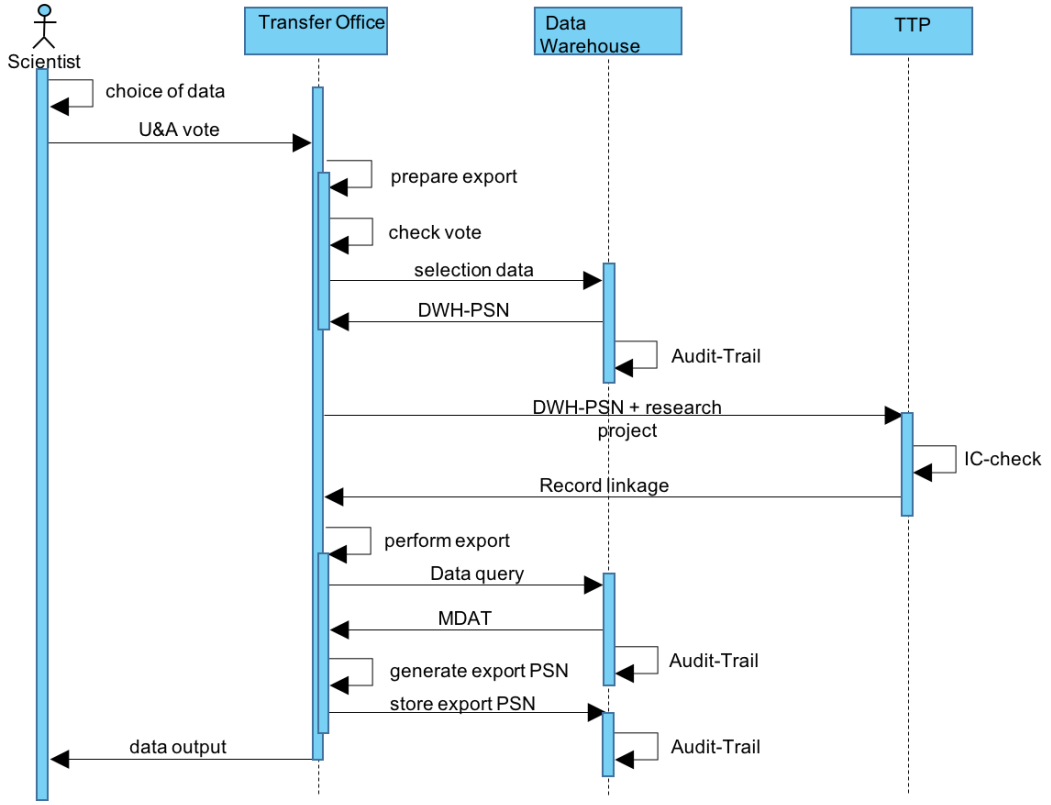


Figure 22: Technical process for retrieving pseudonymised medical data via the Transfer Office

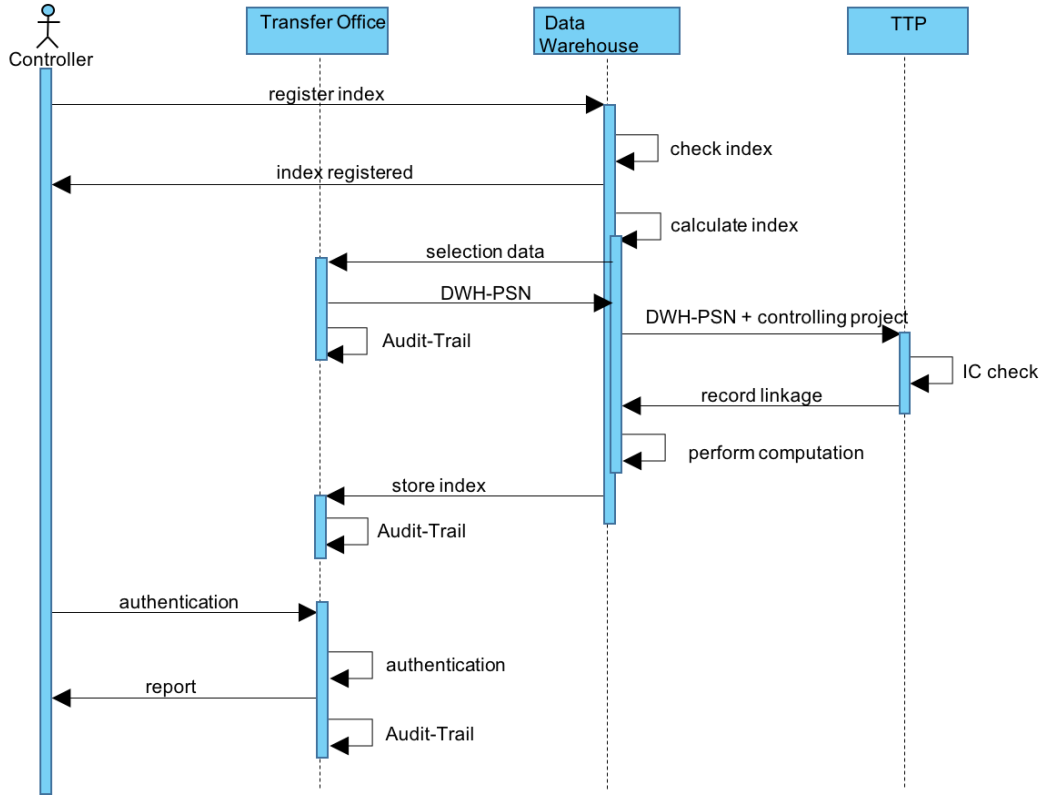


Figure 23: Technical process for retrieving quality management reports via controlling bodies

2 List of abbreviations

IDAT – identifiable information

MDAT – medical data

PSN – pseudonym

MPI– master person index

MPI ID– master person index identifier

IC – informed consent

eCRF – electronic case report form

TTP – Trusted Third Party

LfD MV– Landesbeauftragter für den Datenschutz und Informationsfreiheit Mecklenburg-Vorpommern (State Commissioner for Data Protection and Freedom of Information Mecklenburg-Vorpommern)

3 Glossary

Pseudonymisation (referred to in the official English translation of the German Federal Data Protection Act as "aliasing")– "*[…] replacing a person's name and other identifying characteristics with a label, in order to preclude identification of the data subject or to render such identification substantially difficult."* (BDSG § 3 , [13] [13] [13] [13])

Informed consent – consent, declaration of consent

Master person index – A software system that allows for clear cross-system identification of and differentiation between persons using matching algorithms. Essential for merging personal data from different subsystems.

Homonym error – Data from different persons is erroneously assigned to a single person.

Synonym error– Data from a single person is erroneously assigned several apparently different persons.

Authorisation– The granting of permission to third parties to exercise a right in their own name that they do not normally have.

Consent– An agreement concerning the collection and processing of personal data between the patient and the body responsible for data collection.

Consent– An agreement between the patient, the body responsible for data collection and third parties that regulates the rights to use the collected data, releases the responsible doctor from the duty of confidentiality and allows third parties to perform additional steps using the collected data.

Register– An index for capturing cases (and fatalities) of a certain disease or group of diseases in a defined catchment area (e.g. Germany). A register is numerically complete if all cases in the catchment area have been captured. A register is entirely complete if all the information required for each case has been captured.

4 Bibliography

[1] "Bundesministerium für Bildung und Forschung" ("German Federal Ministry of Education and Research (BMBF)") [Online]. Available at: www.bmbf.de/gesundheitszentren.php. [Accessed: 01 August 2013].

[2] iAS GmbH, "secutrial.com", 2013. [Online]. Available at: http://www.secutrial.com/. [Accessed: 01 August 2013].

[3] C. M. Reng, P. Debold, C. Specker and K. Pommerening, Generische Lösungen zum Datenschutz für die Forschungsnetze in der Medizin [Generic solutions for data protection for research networks in medicine], Medizinisch Wissenschaftliche Verlagsgesellschaft, 2006.

[4] TMF e.V., "tmf-ev.de", 02 July 2013. [Online]. Available at: https://www.tmf-ev.de/Home/Login.aspx?returnurl=/Themen/Projekte/V000_01_PSD/tabid/303/EntryId/2525/Command/Core_Download/Method/attachment/Default.aspx [Accessed: 01 August 2013].

[5] juris.de, "gesetze-im-internet.de", 1990. [Online]. Available at: https://www.gesetze-im-internet.de/bdsg_1990/__3.html [Accessed: 23 January 2013].

[6] M. Lablans, A. Borg and F. Ückert, "unimedizin-mainz.de", 2013. [Online]. Available at: http://www.unimedizin-mainz.de/imbei/medicalinformatics/ag-verbundforschung/mainzelliste.html?L=1 [Accessed: 02 August 2013].

[7] W. Hoffmann, M. Gerlich, C. Schäfer and J. Piegsa, "Datenschutz- und IT-Sicherheitskonzept für die HARMONIC-Studie im Rahmen des HICARE-Verbundprojektes" [Data protection and IT security concept for the HARMONIC study as part of the HICARE joint project], Greifswald, 2013.

[8] T. Hillegeist, Rechtliche Probleme der elektronischen Langzeitarchivierung wissenschaftlicher Primärdaten [Legal problems with long-term electronic archiving of primary scientific data], Göttingen: Universitätsverlag Göttingen, 2012.

[9] German Federal Office for Information Security (BSI), BSI Standard 100-2, Bonn, 2008.

[10] W. Hoffmann, "Rahmenkonzept Datenschutz und IT-Sicherheit für das Institut für Community Medicine der Ernst-Moritz-Arndt-Universität Greifswald (3. überarbeitete Fassung)" [Framework concept for data protection and IT security for the Institute for Community Medicine at the University of Greifswald (3rd Revision)], Greifswald, 2010.

[11] Hessischer Landtag [State Parliament of Hesse], "29. Tätigkeitsbericht des Hessischen Datenschutzbeauftragten" [29th progress report of the State Commissioner for Data Protection, Hesse], Kanzlei des Hessischen Landtags [Chancellery of the State Parliament of Hesse], Wiesbaden, 31 December 2000.

[12] N. Pöttgen, "Medizinische Forschung und Datenschutz, Dissertation" [Medical research and data

protection - a dissertation], in: *Schriften zum deutschen und europäischen öffentlichen Recht* [*Articles on German and European Public Law*], Frankfurt am Main, Peter Lang, Internationaler Verlag der Wissenschaften, 2009.

[13] "German Medicinal Products Act (AMG)", Juris, 2013. [Online]. Available at: https://www.gesetze-im-internet.de/englisch_amg/index.html [Accessed: 01 March 2014].

[14] "Gesetz über Medizinprodukte" [German Medical Devices Act (MPV)], Juris, 2013. [Online]. Available at: http://www.gesetze-im-internet.de/mpg/index.html. [Accessed: 01 March 2014].

[15] "Verordnung über die Anwendung der Guten Klinischen Praxis bei der Durchführung von klinischen Prüfungen mit Arzneimitteln zur Anwendung am Menschen" [Ordinance on the implementation of good clinical practice in the conduct of clinical trials on medicinal products for use in humans], Juris, 2012. [Online]. [Accessed: 01 March 2014].

[16] "Leitfaden Informationssicherheit" [Information Security Guideline], German Federal Office for Information Security (BSI), 2012. [Online]. Available at: https://www.bsi.bund.de/DE/Themen/ITGrundschutz/ITGrundschutzUeberblick/LeitfadenInformationssicherheit/leitfaden_node.html [Accessed: 01 March 2014].

[17] "M 4.223 Integration von Proxy-Servern in das Sicherheitsgateway" [M 4.223 Integration of proxy servers into the security gateway], German Federal Office for Information Security (BSI),2013. [Online]. Available at: https://www.bsi.bund.de/DE/Themen/ITGrundschutz/ITGrundschutzKataloge/Inhalt/_content/m/m04/m04223.html [Accessed: 01 March 2014].

[18] DSK Technology Working Party (DSK Technology Working Party], "Technische und organisatorische Anforderungen: Orientierungshilfe Mandantenfähigkeit" [Technical and organisational requirements: orientation aid and multi-client capability], Arbeitskreises Technik der Konferenz der Datenschutzbeauftragten [Technology Working Party of the German Conference of Commissioners for Data Protection of the Federal Government and the Federal States (DSK)], 2012.

[19] nds-voris.de, "Niedersächsische Vorschrifteninformationssystem" [Legal provision information system for Lower Saxony], juris GmbH, 2012. [Online]. Available at: http://www.nds-voris.de/jportal/portal/t/308u/page/bsvorisprod.psml [Accessed: 01 March 2013].

5 Attachments

I.1 Rahmenkonzept Datenschutz und IT-Sicherheit für das Institut für Community Medicine der Ernst-Moritz-Arndt-Universität Greifswald (Framework concept for data protection and IT security for the Institute for Community Medicine at the University of Greifswald)

I.2 Konzept für den sicheren internen und externen Zugriff auf Forschungsdienste (Concept for secure internal and external access to research services)

Der Landesbeauftragte
für Datenschutz und Informationsfreiheit Mecklenburg-Vorpommern (State Commissioner for Data Protection and Freedom of Information Mecklenburg-Vorpommern)

State Commissioner for Data Protection and Freedom of Information MV
(State Commissioner for Data Protection and Freedom of Information
Mecklenburg-Vorpommern)Lennéstraße 1, Schloss D-19053 Schwerin

REFERENCE
2 . 3. 8 . 038/ 001

YOUR REFERENCE

YOUR LETTER
dated

CONTACT
Werner Baulig
Tel: +49 (0) 385 594 94 46
email: Werner.Baulig@datenschutz-mv.de

18 December 2013

Institute for Community Medicine
Prof. Dr. W. Hoffmann
University Medicine Greifswald
Ellernholzstraße 1-2
D-17487 Greifswald

Trusted Third Party (TTP) data protection concept for the Central Data Management (CDM) joint project / consultation dated 08 November 2013,
Opinion of the LfD MV

Dear Prof. Dr. Hoffmann,

On 08 November, representatives of your institute presented the aforementioned data protection concept, concerning which intense consultation was then provided. The result of this consultation was summarised in a protocol dated 26 November 2013. This protocol requires no significant changes or additions on our part.

In our assessment and on the basis of the expert suggestions recorded in this protocol, the presented data protection concept complies with applicable data protection law and is supported by us.

We would like to make the following recommendations to optimise its implementation:

1. Implementation of the TTP in a university statute
2. Maximum consideration of the DSK's orientation aid entitled "Mandatenfähigkeit" (Multi-client capability) dated 2012(https://www.datenschutz-bayern.de/technik/orient/oh_mandantenfaehigkeit.pdf).

We would be grateful if you could send us an updated version of the concept including the results of the protocol at your convenience.

Finally, we would like to wish you a peaceful Christmas and good health for the new year in 2014. We would like to take this opportunity to thank you and your staff again for your extremely proactive and constructive professional cooperation to achieve successful data protection.

Kind regards,

p.p.

Werner Baulig

Der Landesbeauftragte für den Datenschutz Niedersachsen (State Commissioner for Data Protection Lower Saxony)

Der Landesbeauftragte für den Datenschutz Niedersachsen (State Commissioner for Data Protection, Lower Saxony) Postfach 2 21• D-30002 Hanover

University Medical Centre
GöttingenDepartment of Medical Informatics
Mr. Quade
Robert-Koch-Str. 40
D-37075 Göttingen

Processed by

Mr Kraul

Your reference, your letter dated	**My reference** (please quote in correspondence)	**Extension+49 (0)511 120-**	**Hanover,**
13 May 2014	2.2 - 1759-316	45 12	19 June 2014

Data protection concept for the Data Handling unit of the Central Data Management at the German Centre for Cardiovascular Research (DZHK)

Dear Mr Quade,

In your letter dated 13 May 2014 you sent me Version 1.2 dated 24 March 2014 of the process description and data protection concept of the DZHK Central Data Management unit and the opinion of your data protection officer Dr. Döler, dated 29 April 2014, with the request to assess the documents with regard to legal aspects of data protection.

Based on the documents presented to me, the telephone conversation with Dr. Döler on 27 May 2014, your response to my queries dated 06 June 2014 and a conversation with my colleague in Mecklenburg-Vorpommern, I am pleased to inform you that the present data protection concept by the Data Handling unit of the DZHK Central Data Management unit complies with applicable data protection law.

I wish you every success with this project.

Yours sincerely,p.p.

Kraul [

D-30159 Hanover

Visitors
Prinzenstr. 5

Tel:
+49 (0) 511 120 4500

Fax:
+49 (0) 511 120 4599

Internet:
www.lfd.niedersachsen.de
email:
poststelle@lfd.niedersachsen.de