Template for the Data Protection Concept of an   
OSSE Patient Registry

Dr. Marita Muscholl\*\*, Martin Lablans\*\*, Andreas Borg\*\*, Prof. Dr. Frank Ückert\*\*, Prof. Dr. TOF Wagner\*

\*Frankfurt University Hospital

Dept. of Pulmonology

69590 Frankfurt am Main, Germany

\*\*Johannes Gutenberg University Medical Center Mainz

Institute of Medical Biostatistics, Epidemiology, and Informatics (IMBEI)

55101 Mainz, Germany

Notes on Using this Template

The goal of this template for the data protection concept of an OSSE registry is to facilitate the development of an individual data protection concept for the implementation of the OSSE registry software. The template considers all components that are part of the OSSE registry software or are used in its context, along with the data flows between these components. Furthermore, it describes data protection measures that are part of the software features. Information to be complemented on organizational processes or on further IT components used is marked in the texts. Technical measures not part of the software should be considered suggestions.

***Please note:*** *This document is a translation of the original German data protection template for OSSE registries based on German data protection law and regulations as well as published guidelines on the topic. When using this template, please be aware that you must identify and take into account the appropriate legal and regulatory requirements.*

### Markings and their meaning

* *Passages in the text where information needs to be complemented or place holders to be substituted are marked gray* [*and put into square brackets as comments where necessary*]
* The beginnings and endings of those sections in the text that require different versions depending on the local conditions are identified by yellow markings in square brackets:  
  [*#beginning* ***version n*** *(version is described here)*] … [*#end* ***version n***]  
  [*Only applies to* ***version m***] …
* Notes on particular sections are also put in square brackets and marked yellow: [*This section should be edited in the following way…*]
* CAUTION: Modifications to the text outside the yellow areas might change the concept with regards to privacy laws and content. Deviations from the underlying TMF data protection concept[[1]](#footnote-2) will very likely lead to a loss of the basic approval by all German state data protection agencies.

### Adaptation Procedure

1. Substitute the place holder “OSSE Registry for Rare Disease X” with the name of the planned registry throughout the template.
2. Answer the following questions for yourself, which are necessary for the choice of versions or information to be complemented.
3. *Who is the governing body (i.e., usually the “owner”) of the registry?*
4. *Which institution runs the OSSE registry (i.e. usually controls the necessary central hardware)?*
5. *Which institutions are participating in the project (i.e. record data into the registry)?*
6. *Which data are being recorded into the OSSE registry (e.g. weight, date of birth, diagnosis as Orphacode)? (🡪 will result in an appendix)*
7. *Where is the identity management for the pseudonymization of the patient data located and who is running it? (The following options are possible:)*
   1. *Identity management is handled centrally for all registries for rare disease x (so that, in research projects involving other registries, identical patients can be recognized in various registries)[[2]](#footnote-3).*
   2. *The OSSE Registry for Rare Disease X shall involve a proprietary identity management system.*
   3. *No identity management component is used. The OSSE registry is configured in a way that it issues internal patient identifiers (PIDs) that are linked manually with the patient’s identifying data exclusively at the place of treatment.*
8. *Which roles are assigned within the registry (e.g. “medical data manager”, “treating doctor”)? Which respective access rights will they have to be equipped with (e.g. “reading of xyz”, “writing of abc”)?*
9. *What does the process of manual data entry look like (which states, e.g. “in process”, “completed”, “validated”, can an entry form have and which user groups [roles] are allowed to work on it)?*
10. *Are data being imported from clinical systems?*

*If yes, which data from which systems? (🡪 will result in an appendix)*

1. *Are data on biomaterial samples imported into the OSSE registry (and which)?*

*If yes, □ from a biobank? □ directly from the sample source?*

1. Substitute the placeholders and complement the requested information and appendices.
2. Erase all unnecessary sections.
3. Read through the template carefully and adjust the text concerning your specific environment (IT infrastructure, organization etc.).
4. Erase the comments and markings.
5. Update the table of contents.

# Logo_UKF_-_rgb



OSSE Registry for Rare Disease X

Data Protection Concept[[3]](#footnote-4)

Project OSSE Registry for Rare Disease X

Authors [*Author(s)*]

[*Institution(s)*]

[*Governing Body (Name and Address)*]

Version [Click to set date]

Content

[1. Introduction 6](#_Toc415043261)

[1.1 Objective 6](#_Toc415043262)

[1.2 Data Processing Overview 6](#_Toc415043263)

[1.3 Legal Basis 8](#_Toc415043264)

[1.4 Governing Body 8](#_Toc415043265)

[2. Data Processing Components 8](#_Toc415043266)

[2.1 OSSE Registry 8](#_Toc415043267)

[Components and Features 8](#_Toc415043268)

[Workflow 9](#_Toc415043269)

[Users, Roles, and Rights 9](#_Toc415043270)

[2.2 Identity Management 9](#_Toc415043271)

[Pseudonyms 9](#_Toc415043272)

[Manual Linking 10](#_Toc415043273)

[2.3 Biobank 10](#_Toc415043274)

[2.4 OSSE Share Client 10](#_Toc415043275)

[2.5 Metadata Repository 10](#_Toc415043276)

[2.6 Registry of Registries 11](#_Toc415043277)

[3. Data Processing Procedures 11](#_Toc415043278)

[3.1 Manual Data Entry 11](#_Toc415043279)

[Creating a Patient Record 11](#_Toc415043280)

[Selection of Patients, Data Entry 11](#_Toc415043281)

[Workflow Support 11](#_Toc415043282)

[Manual Sharing of Data Sets 11](#_Toc415043283)

[3.2 Data Import 11](#_Toc415043284)

[3.3 Pseudonymization 12](#_Toc415043285)

[Manual Patient Registration 12](#_Toc415043286)

[Pseudonymization at Data Import 12](#_Toc415043287)

[Creating and Managing Keys 13](#_Toc415043288)

[3.4 Importing Biomaterial Data 13](#_Toc415043289)

[3.5 Data Export 16](#_Toc415043290)

[3.6 Decentralized Search 17](#_Toc415043291)

[3.7 Uploading Information to the Registry of Registries 17](#_Toc415043292)

[4. Organizational Framework 17](#_Toc415043293)

[4.1 Operation of Components 17](#_Toc415043294)

[4.2 Participating Researchers 17](#_Toc415043295)

[4.3 Members’ Meeting 18](#_Toc415043296)

[4.4 Registry Board 18](#_Toc415043297)

[4.5 Registry Commission 18](#_Toc415043298)

[4.6 Access by System Administrators 18](#_Toc415043299)

[5. Data Protection Provisions 19](#_Toc415043300)

[5.1 Informational Separation of Powers 19](#_Toc415043301)

[5.2 Authorization and Authentication 19](#_Toc415043302)

[User Authorization 19](#_Toc415043303)

[Authorization of Components 19](#_Toc415043304)

[User Authentication 19](#_Toc415043305)

[Authentication of Components 19](#_Toc415043306)

[5.3 IT Infrastructure Provisions 19](#_Toc415043307)

[Security of Stored Data 19](#_Toc415043308)

[Security of Communication 19](#_Toc415043309)

[Logging 20](#_Toc415043310)

[6. Observing the Rights of Affected Individuals 20](#_Toc415043311)

[6.1 Information and Consent 20](#_Toc415043312)

[6.2 Information on Stored Data 20](#_Toc415043313)

[6.3 Revocation, Deletion, Anonymization 21](#_Toc415043314)

[6.4 Storage Period 21](#_Toc415043315)

[7. Appendix 22](#_Toc415043316)

[7.1 OSSE Registry Patient Informed Consent 22](#_Toc415043317)

[7.2 Data Sets 23](#_Toc415043318)

[Registry Definition 23](#_Toc415043319)

[Identifying Data 23](#_Toc415043320)

[7.3 Data and Source Systems for Data Import 24](#_Toc415043321)

# Introduction

## Objective

[*The following section is a suggestion for phrasing and should be substituted or adjusted and specified.*]

Given the lack of knowledge on rare disease x, the medical and social responsibility for the persons affected by it calls for the monitoring of the progression over long periods of time.

Sensible clinical and basic research into rare diseases such as rare disease x is only possible in multi-location networks with sufficient case numbers. Also, reliable information on the incidence of certain manifestation patterns, health status, etc. is of utmost importance for health care and health policy.

The OSSE Registry for Rare Disease X is committed to help improve the research conditions by consolidating this kind of information in one place.

To this end, the OSSE Registry for Rare Disease X can be linked to other Registries for Rare Disease X in an IT infrastructure that particularly facilitates the determination of case numbers across registries, but also the recruitment of subjects for clinical studies or the exchange of available data for evaluations to be used in specific research questions. For this purpose, participating registries collect data from clinical documentation and on available biomaterial samples. To make these data usable for research, the OSSE Registry for Rare Disease X provides a search interface which makes it possible to

* evaluate whether a sufficient number of subjects with the characteristics required for a given research project is available in the OSSE Registry for Rare Disease X,
* determine which institutions are currently treating or have treated suitable patients, and
* make inquiries concerning the use of these patients’ medical data and biomaterial samples for research purposes.

In doing so, the OSSE Registry for Rare Disease X ensures that data sovereignty remains with the persons in charge of the registry and that data can only be exchanged in a way that precludes the identification of individual patients by choosing appropriate pseudonyms.

## Data Processing Overview

The OSSE Registry for Rare Disease X collects and processes data of patients treated in the participating hospitals (also called “locations”). The data are recorded in the following way:

[*Delete where inapplicable!*]

* manually via the web-based OSSE user interface
* via data import from existing data processing systems, e.g. from hospital information and documentation systems (see 7.3, “Data and Source Systems for Data Import”)

The following data are collected:

* Identifying data (IDAT): this includes demographic data (e.g. name, date of birth, sex) that allow for the definitive identification of the patient. They are not stored in the registry but in a separate patient list.
* Medical data (MDAT): this includes all data on the disease and its course that are recorded in the registry.
* Data on biomaterial samples the OSSE Registry for Rare Disease X also contains data on biomaterial (see 3.4, “Importing Biomaterial Data”).

More detailed information on the data scope can be found in the appendix under 7.2, “Data Sets”.

All data fields stored in the registry are registered and described in a metadata repository operated centrally for all registries. The registry structure (master and longitudinal data forms) is defined dynamically by means of a registry editor (see section 7.2, “Data Sets” – “Registry Definition”).

[*#Beginning* v***ersion 1****: Centralized ID management for all Registries for Rare Disease X*]

IDAT are stored in a centralized ID management system run by [*specify the institution running the* Mainzelliste*[[4]](#footnote-5) centrally in trust for all registries for rare disease x*].

[*#End version 1*]

[*#Beginning* ***version 2****: Centralized ID management for a specific Registry for Rare Disease X*]

IDAT are stored in OSSE Registry for Rare Disease X’s ID management system run by [*specify the institution running the* Mainzelliste *in trust for this registry*]

[*#End version 2*]

[*#Beginning* v***ersion 3****: Manually managed list of IDAT; in this case, OSSE creates a PID registered in the list together with the IDAT*]

The OSSE registry produces a pseudonym when a patient record is created. IDAT are managed manually locally at the registry locations and assigned to the pseudonym.

[*#End version 3*]

Data in the OSSE Registry for Rare Disease X are exported for analysis with non-traceable export pseudonyms. The OSSE Registry for Rare Disease X further supports “decentralized search,” an infrastructure for inquiries that allows networked OSSE registries to be searched for specified cases. The inquiry, consisting of an exposé of the question, the inquirer’s contact information, and the search criteria, is then automatically presented electronically to the registry management together with the results. The management then has the option to check the inquiry for content criteria and legal legitimacy and, provided both are positive, to manually answer the inquiry and transmit the data sets if applicable. This method comes closest to a written inquiry, except that due to the cross-linking of registries, it is already clear where an inquiry concerning suitable data sets can make sense, and that the assessment and response are supported by technical resources. To facilitate comparability and targeted inquiries for data content in different data sources, the data schema of an OSSE registry is linked to a formal definition of all data fields registered as metadata in the central metadata repository (MDR).

Sections 2 and 3 provide a detailed description of the components and processes.

## Legal Basis

The patient’s informed consent (see section 6.1, “Information and Consent”) constitutes the legal basis of data processing. It explicitly mentions the institutions and persons allowed to process and use specified data. It also considers the sharing of data for research purposes which are quasi-“anonymized” via a non-traceable export pseudonym, since particularly in the area of rare diseases one cannot preclude the possibility of tracing the patient’s identity from the medical data. An inquiry to networked OSSE Registries for Rare Disease X via “decentralized search” is not relevant concerning data protection provisions, as it only yields aggregated data (e.g. number of specified cases) and results are not transmitted automatically. Data transfer requires consent.

## Governing Body

The governing body is [specify the governing body here].

# Data Processing Components

## OSSE Registry

### Components and Features

The OSSE registry software serves the recording and storage of master and longitudinal medical data of patients affected by rare disease x. Data fields and forms are externally registered and described in a metadata repository or form repository, respectively, and can be modified and supplemented even after the beginning of the project.

Data entry is done at the individual locations via the web-based user interface; additionally, data can also be collected via data import interfaces. All data are stored in a versioned way, i.e. changed and deleted values remain available in the database and can be shown if necessary. [[5]](#footnote-6)

[*#Beginning* ***version 1, version 2*** *(Identity management with central or local component as defined in section 2.2):*]

Identifying data are not recorded into the OSSE registry but directly into the identity management system. Communication between the identity management and the OSSE registry happens via a web browser. The user sees the ID management’s entry mask integrated into the OSSE registries’ user interface. The returned pseudonym, “PSNOSSE“ (see section 2.2, “Identity Management “), is stored with the MDAT, but not shown, so that is impossible to correlate IDAT and PSNOSSE outside the ID management system even manually. However, since medical data are recorded locally close to the time of treatment, IDAT and MDAT can be shown in the browser together. This is done by way of temporary identifiers via which the browser retrieves e.g. a patient’s first and last name and which ensure that the correlation between PSNOSSE and the patient’s IDAT does not become known outside the ID management.

[*#End version 1, version 2*]

[*#Beginning* ***version 3*** *(Identity management by manually managed list)*] Identifying data are not recorded into the OSSE registry. To mark the data set, the OSSE registry shows the identifiers created internally (PIDOSSE). The correlation between IDAT and PIDOSSE is done manually at the registry location.

[*#End version 3*]

### Workflow

Medical data are recorded into forms (master and longitudinal data forms) that can take on the following status values during data processing:

[*Describe the document status values and status transitions here if applicable; if not, delete the whole section.*]

### Users, Roles, and Rights

Access rights are granted based on roles. Depending on his/her position, each user has one or several roles with which he/she logs in. To define access rights, data are classified particularly concerning their attribution to the informational unit in which they were collected [*and concerning… (add further criteria here)*]. The following roles are available for OSSE Registry for Rare Disease X according to the declaration of consent, participating locations, and organizational requirements:

[*List of roles and their descriptions/assigned rights; as appendix if necessary*]

Each access is recorded (see section 5.3 “IT Infrastructure Provisions”).

## Identity Management

Pseudonymization is a necessary measure to keep a high level of data protection in order to protect the patient from reverse identification. His/her identifying data (IDAT) are substituted by pseudonyms. When a pseudonym is requested, the data set is checked for correspondence with existing data sets (record linkage). Depending on the degree of IDAT correspondence and on the threshold values set, a new data set is created or an existing one returned.

### Pseudonyms

[*#Beginning* ***version 1*** *(Centralized ID management for all Registries for Rare Disease X):*]

For pseudonymization, the OSSE Registry for Rare Disease X uses an instance of the *Mainzelliste* that is managed centrally for all registries for Rare Disease X by [manager of the central *Mainzelliste*]. It creates a unique identifier (PID) for each patient as well as one independent second-level pseudonym (PSN*OSSE*(#)) for each OSSE registry which collects the patient’s data. The *Mainzelliste* also creates independent pseudonyms for the following research systems in the Registry for Rare Disease X:

* [List of further systems within the research network storing pseudonymized data and type of pseudonyms]

Furthermore, the *Mainzelliste* creates non-traceable export pseudonyms for the export of patient data for research purposes, with data of identical patients receiving the same export pseudonym if cases from different Registries for Rare Disease X are combined. The centralized ID management allows for a patient’s patient-relatable data collected or imported on different occasions at/from different locations (and into different registries) to be combined according to data protection rules.

[*#End version 1*]

[*#Beginning* ***version 2*** *(Local ID management for this Registry for Rare Disease X)*]

For psydonmyization, OSSE Registry for Rare Disease X uses an instance of the *Mainzelliste* run by [manager of the central Mainzelliste]. It creates a unique identifier (PID) and a second-level pseudonym (PSN*OSSE*(#)) for each patient. The *Mainzelliste* also creates independent pseudonyms for the following research systems in the Registry for Rare Disease X:

* [List of further systems within the research network storing pseudonymized data and type of pseudonyms]

Furthermore, the *Mainzelliste* creates non-traceable export pseudonyms for the export of patient data for research purposes, with data of identical patients receiving different export pseudonyms if cases from different registries or bridgeheads for rare disease x are combined. This means that the local ID management for the OSSE Registry for Rare Disease X allows for a patient’s patient-relatable data collected or imported on different occasions at/from different locations (and into different registries) to be combined according to data protection rules, but data on identical patients from different registries cannot be correlated.

[*#End version 2*]

[*#Beginning* ***version 3*** *(Manually managed list of IDAT; in this case OSSE creates a PID that is recorded in the list together with the IDAT)*]

When creating a patient record, OSSE Registry for Rare Disease X produces an internal identifier (PIDOSSE). The PID is managed in the treatment context **on location** in a list together with the IDAT.

[*#End version 3*]

### Manual Linking [*Only for version 1 and version 2*]

An interface allows a person to check and, if necessary, correct the results of automated matching, i.e. to merge duplicates or separate falsely merged data sets. To do so, the match weights (reference values to compare the individual attributes of patients to be checked) are shown and it is possible to draw on medical data to make a decision.

## Biobank

[*In case data are imported from a biobank: include here the description of the biobank, its manager, data stored, etc.; as appendix if necessary*]

## OSSE Share Client

The share client is the client interface for decentralized search (see section 3.6, “Decentralized Search”). It fetches inquiries placed via the search broker, communicates with the OSSE store, and shows the authorized user (e.g. a member of the registry board) inquiries and results. In doing so, only pseudonymized and aggregated data are shown. The data manager can launch a data export with the share client interface. The exported data contain non-traceable export pseudonyms instead of the pseudonyms stored in OSSE. The decision about the transfer of such “quasi-anonymized” data to third parties lies with [*specify the body in charge here*]. The transfer of the exported data must conform to the data uses described in the informed consent, since particularly in the area of rare diseases there is a high risk of re-identification through medical data alone.

## Metadata Repository

The Metadata Repository (MDR) stores the meaning (semantics) of all (reference) data elements used in the OSSE Registry for Rare Disease X. It offers a controlled vocabulary (syntax) and can provide machine-readable, structured information on data elements, e.g. conceptual domains or value ranges. Furthermore, it defines the fields of the registry forms specified in this concept (see section 7.2, “Data Sets”). Since the MDR does not process personal data, it will not be treated in more detail here.

## Registry of Registries

All OSSE registries and bridgeheads[[6]](#footnote-7) register in the registry of registries with a short description of the registry, contact persons and, if applicable, additional registry metadata and interesting core data (e.g. overall number of cases). The data are uploaded actively by the persons in charge. No patient data are transmitted. The registry of registries is managed by [enter the manager of the registry of registries here].

# Data Processing Procedures

## Manual Data Entry

### Creating a Patient Record

1. The user enters the IDAT into the *Mainzelliste* entry form integrated into the OSSE registry.
2. The OSSE registry receives a pseudonym (see section 3.3, “Pseudonymization” for further detail) that is stored together with the data set. The user does not receive feedback on whether the patient record is already present in the *Mainzelliste* or a new record was created.

### Selection of Patients, Data Entry

The user receives a list of patients (with real names) depending on his/her access rights. After selecting a patient, he/she can edit forms, and the browser window shows IDAT of the selected patient.

### Workflow Support

[*Describe the data recording process here: Which status does a document have when? Who is allowed to see and edit documents?]*

### Manual Sharing of Data Sets

In justified cases, e.g. when requesting second opinions, data sets can be shared with defined persons and roles. The share includes individual forms or complete cases (all forms of a patient at the respective location). The user performing the share must be explicitly authorized to do so. A share is valid only for a limited period. The user selects a patient data set for sharing. In the dialog window “share data set”, he/she determines the user or role to be authorized and the validity period. The data share has to be covered by the access rights and purposes described in the patient informed consent.

[*Please delete the following section if no data are imported from other systems.*]

## Data Import

Data from existing systems are taken over into the OSSE Registry for Rare Disease X at regular intervals and created or updated there. An overview of data to be imported and their source systems is available in Appendix 7.3, “Data and Source Systems for Data Import”. The data import happens in the following stages of a so-called ETL[[7]](#footnote-8) process, with the data always remaining at the respective locations’ servers until they are imported into the OSSE registry:

1. Identifying, medical, and sample data are extracted from source systems.
2. Identifying data are substituted by a pseudonym within the transformation step (see section 3.3, “Pseudonymization” for further detail)
3. The ETL process loads the data into the OSSE registry via a web interface.

## Pseudonymization

[*#Beginning* ***version 1*** *and* ***version 2*** *(local or central* Mainzelliste*)*]

Pseudonymization is part of any kind of data collection into the OSSE registry – manual data entry and automated data import.

### Manual Patient Registration

For both the registration of a new and the retrieval of an existing patient data set, the user enters identifying data into the *Mainzelliste* entry form displayed in the OSSE GUI browser window. The identifying data have to be entered completely, since drop down lists familiar e.g. from clinical workspace systems cannot be displayed here upon entering name parts. A record algorithm checks whether the patient has already been registered in the *Mainzelliste*. If not, a new patient record is created by storing the IDAT and producing a non-speaking PID as well as the PSNOSSE as a second-level pseudonym. The user is then automatically redirected to a website of the OSSE registry where MDAT for the newly created or selected patient record can be entered. In doing so, the browser and the OSSE registry communicate by way of temporary identifiers. The PSNOSSE does not become visible for the user, i.e. it does not appear in the HTML code of the forms displayed or in the browser’s HTTP inquiries either. This procedure ensures that PSNOSSE and IDAT cannot be correlated outside the *Mainzelliste* at any point.

During MDAT entry into the OSSE registry, which takes place locally and close to the time of treatment, the patient’s identifying data are shown in the browser. These are correlated to the MDAT only in the browser, so that the OSSE registry does not get access to IDAT at any point. For that purpose, the registry software retrieves a session-based temporary ID for each PSNOSSE from the *Mainzelliste*, with which the browser then receives the corresponding IDAT from the *Mainzelliste*.

The pseudonyms stored in the OSSE registry are never displayed or released, so that they cannot be assigned to a patient either at data entry in the treatment context or by merging exported data. Re-identification (de-pseudonymization) can only be performed in a controlled manner by means of the *Mainzelliste*.

[*#End version 1, version 2*]

[*#Beginning* ***version 3*** *(OSSE creates its own PIDs)*]

Upon creating a new patient record, the registry produces a unique patient identifier (PID). PID and IDAT are correlated locally at the location of treatment in a manually managed list as follows:

[*Describe here how the process is organized and how the list is protected from access by unauthorized persons.*]

[*#End version 3*]

### Pseudonymization at Data Import

Upon data import from [the clinical systems or biobank/sample source (*delete as applicable!)*], IDAT are substituted by pseudonyms before the data sets are loaded (see section 3.2, “Data Import “). Data extraction and transformation are supported by data integration software (Talend Open Studio). The pseudonymization is part of the data transformation and carried out with a component developed particularly for this purpose. The process consists of the following steps:

1. For each data set, the data integration software launches the *Mainzelliste* and turns over the IDAT.
2. The *Mainzelliste* determines or creates the PSNOSSE (analogous to manual data entry) and encrypts it with a public key of the OSSE registry so that the OSSE pseudonym cannot be correlated with the IDAT outside the *Mainzelliste*.
3. The *Mainzelliste* sends the encrypted pseudonym, (PSNOSSE)tr, to the transformation component, where the IDAT are substituted by the (PSNOSSE)tr.
4. The import interface decrypts the (PSNOSSE)tr and stores the data with the PSNOSSE.

This procedure ensures that pseudonyms cannot be correlated during data import either, since the data-providing party providing, which could correlate them to patients, can only see the encrypted (PSNOSSE)tr.

### Creating and Managing Keys

The pair of keys for the protected transfer of OSSE pseudonyms is created in the OSSE registry upon startup and stored during runtime. A new pair of keys can be produced upon re-start or through a registry feature. The current public key can be retrieved from the OSSE registry via a web interface at all times by way of an authorized component (e.g. ID management).

[*The following section (“Importing Biomaterial Data”) is required if biomaterial samples are supposed to be imported into the registry. It has to be distinguished here whether the sample data originate directly from the sample source or whether the research network uses a separate biobank. The latter case presumes that the biobank, according to the TMF Data Protection Concept for Biobanks[[8]](#footnote-9), does not pass on the LabID from the sample source in clear text but only in encrypted form and does not store any patient correlation itself. Differing implementations have to be adjusted accordingly in the following section.*]

## Importing Biomaterial Data[[9]](#footnote-10)

[*#Beginning biobank exists*]

The link to a biobank should provide information on biomaterial samples facilitating the identification of those patients in the registry who are suitable for a particular research project. Data on biomaterial samples (IDAT, PID, LabID, further characteristics of the sample) are recorded in the treatment context (sample source) and transmitted to the biobank. The sample’s unique identifier, LabID, created at the sample source, is encrypted in the biobank into (LabID)tr. The samples stored in the biobank are not patient-related; the samples’ link to the patient for the transfer to the registry should therefore only be stored temporarily (until sample data are transferred). The sample data in the registry do not allow for information on the sample to be retrieved directly from the sample source. The following data are imported into the OSSE registry from the biobank:

[*Delete items from this list as applicable!*]

* Existence and/or number of samples, respectively
* Sample characteristics
* (LabID)tr

The import of biomaterial data consists of the following steps:

1. The patient is registered in the *Mainzelliste* via the sample source: the sample source sends IDAT and receives the encrypted OSSE pseudonym (PSNOSSE)tr and, if necessary, the universal PID.
2. The patient record is created in the OSSE registry by a user
3. The sample is sent to the biobank: the address label contains the (PSNOSSE)tr, which is buffered for the data transfer to the registry.
4. The sample is registered in the biobank module: the biobank module stores the LabID and all necessary sample data (OrgDat). Through a cryptographic process, the LabID becomes the LabIDtr in order to avoid direct correlation of MDAT and sample. The correlation of (PSNOSSE)tr and LabIDtr is stored temporarily.
5. Transfer of sample data to the OSSE registry: For each sample, the OSSE registry receives a data set consisting of (PSNOSSE)tr, LabIDtr and further information on the sample that is supposed to be transmitted as well (see above). As described in section 3.2, “Data Import“, the data are transformed and loaded into the OSSE registry. After successful transfer, the link between (PSNOSSE)tr and LabIDtr is deleted. This avoids a permanent correlation between sample and patient in the biobank.

Figure 1: Data flows during biomaterial data import

[#*End biobank exists*]

[*#Beginning biobank does not exist*]

The OSSE Registry for Rare Disease X should provide information on biomaterial samples facilitating the identification of those patients in the registry that are suitable for a particular research project. Data on biomaterial samples (IDAT, PID, LabID, further characteristics of the sample) are recorded in the treatment context (sample source).

Figure 1: Data flows during biomaterial data import

The following data are imported into the OSSE registry:

[*Delete items from this list as applicable!*]

* Existence and/or number of samples, respectively
* Sample characteristics

The import of biomaterial data consists of the following steps:

1. The patient is registered in the *Mainzelliste* via the sample source: the sample source sends IDAT and receives the encrypted OSSE pseudonym (PSNOSSE)tr and, if necessary, the universal PID.
2. The patient record is created in the OSSE registry by a user
3. Transfer of sample data from the sample source to the OSSE registry: For each sample, the OSSE registry receives a data set consisting of (PSNOSSE)tr, and further information on the sample that is supposed to be transmitted as well (see above). As described in section 3.2, “Data Import“, the data are transformed and loaded into the OSSE registry.

[In case storing the *(PSNOSSE)tr* in the sample source temporarily is not possible, it can be retrieved through the IDAT via the pseudonymization step in the ETL. *The section needs to be adjusted accordingly.*]

[*#End biobank does not exist*]

## Data Export

Data can be exported for analysis. For this purpose, the OSSE registry’s internal pseudonym is substituted by an export pseudonym during export.

[*#Beginning* ***version 1*** *(centralized ID management)*]

If data from different registries are brought together in the context of decentralized search, uniform non-traceable export pseudonyms are queried with the centralized ID management upon export. These facilitate the matching of data sets from identical patients recorded in different registries, as well as consistent updates of the combined data sets. The export takes place in the following steps:

1. OSSE sends the internal pseudonym PSNOSSE(#) to the ID management along with a project identifier.
2. The ID management returns a uniform project-specific export pseudonym (PSNProjekt).
3. The data set is exported with the PSNProjekt.

[*#End version 1*]

[*#Beginning* ***version 2*** *(Local ID management)*]

If data from different registries are brought together in the context of decentralized search, uniform non-traceable export pseudonyms are queried with the local ID management upon export. These facilitate consistent updates of the combined data sets, but identical patients from different registries are not recognized. The export takes place in the following steps:

1. OSSE sends the internal pseudonym PSNOSSE(#) to the ID management along with a project identifier.
2. The ID management returns a project-specific export pseudonym (PSNProjekt).
3. The data set is exported with the PSNProjekt.

[*#End version 2*]

Particularly in diseases with low case numbers, if the course of disease or additional data are known, medical data can be related to specific patients even without knowing the identifying data. Even the use of non-traceable export pseudonyms cannot always guarantee factual anonymity.

For this reason, the export of pseudonymized data and the purpose of their use are considered in the patient informed consent.

## Decentralized Search

Through decentralized search, researchers can search the databases of OSSE registries or bridgeheads in order to find registries containing data and samples potentially relevant to a research project. The search broker provides a web-based search form to record the inquiries that filter data elements by preset values or free-text search. Several search attributes can be combined at will via logical operators. The search form also transmits an abstract of the intended research project and the inquiring researcher’s contact information.

In a first step, the inquiry is saved and the inquiring researcher receives a corresponding notification. The participating registries’ share clients fetch new inquiries from the search broker at regular intervals and determine which data sets in the OSSE registry match the search criteria. For each registry, an authorized person can view the inquiry’s content and the data sets found and subsequently contact the inquiring researcher to arrange a potential data or sample transfer. This process and the data protection issues associated with it, e.g. the necessity to obtain further consent, have to be clarified for each individual case by the participating parties. Data transfer happens in a controlled way outside the OSSE registry.

## Uploading Information to the Registry of Registries

Information on the OSSE Registry for Rare Disease X (new registrations and updates) can be uploaded to the registry of registries through a menu feature executed by an authorized user. The scope of information to be uploaded can be set in the registry configuration. This does not apply to registry content data (see section 2.6 “Registry of Registries”).

# Organizational Framework

[*Along with the organizational framework, specify the governing body’s stipulations (e.g. bylaws or similar) legally reliable in case of a lawsuit here.*]

## Operation of Components

The OSSE Registry for Rare Disease X is operated by [*insert location/institution running and managing the OSSE registry here*]. Data recording locations for the registry include:

* [*Insert list of institutions/locations here*]

Operation of the OSSE Registry for Rare Disease X’s central components is managed by selected institutions appointed upon agreement with the registry’s steering committee (registry board). For the purpose of the informational separation of powers, the organizational framework ensures independent operation of ID management and OSSE registry (see section 5.1, “Informational Separation of Powers”).

## Participating Researchers

*Participating Researchers* are individuals who can place inquiries via decentralized search. Generally, all members of the registry locations can use OSSE Registry for Rare Disease X as participating researchers, with each location deciding by itself which of its members are granted access (see also section 5.2, „Authorization and Authentication“.

Scientists who are not a member of an OSSE Registry for Rare Disease X can request access from the registry commission (see also section 4.5, “Registry Commission”). Access authorization should be adequately limited in time.

## Members’ Meeting

Participating researchers/clinicians (or their institutions, respectively) sharing their data in the OSSE Registry for Rare Disease X participate in the registry’s members’ meeting with one vote.

## Registry Board

The registry board is elected by the OSSE Registry for Rare Disease X’s members’ meeting to manage the registry’s business.

## Registry Commission

The OSSE Registry for Rare Disease X’s registry board appoints a registry commission in charge of the following duties, among others:

* Review and approval of requests to use the OSSE Registry for Rare Disease X by external researchers[[10]](#footnote-11) (decentralized search)
* Review and approval of requests to export medical data for external research projects
* Review and approval of requests to notify affected patients of research results

Furthermore, the registry commission is the first point of contact for data protection issues.

The registry commission is staffed in a way that each of the OSSE Registry for Rare Disease X’s locations is represented. At the least, members include:

* A doctor mainly working with affected patients
* A scientist researching with the data administered in the OSSE Registry for Rare Disease X (or similar data)
* A data protection officer or a lawyer familiar with the subject of data protection

A representative of the OSSE Registry for Rare Disease X’s developer team can be consulted to provide advisory support.

## Access by System Administrators

Generally, the data stored in OSSE Registry for Rare Disease X can be viewed by the administrators of the IT infrastructure used. Administrators may only access the data if it is essential to performing their duties. The data access procedure is regulated as follows: [*Describe here how such access takes place and how it is documented, e.g. beyond the usual logging and indicating the reasons*]. All administrators have to be instructed accordingly and agree to maintain confidentiality[[11]](#footnote-12).

# Data Protection Provisions

## Informational Separation of Powers

ID management is operated separately (logically, physically, and organizationally) from all components storing MDAT or data on biomaterial samples. [*The person/institution in charge of ID management (specify the person/institution in charge of ID management here)*] operates at his/her/its own legal responsibility and is not subject to the directives of the registry management. This ensures that individuals with access to clinical or biomaterial data in the OSSE Registry for Rare Disease X outside the treatment context are not able to correlate the data to real patients.

## Authorization and Authentication

### User Authorization

User authorization (assigning defined roles to users) in the OSSE registry is done through local administrators at the respective locations according to the local structures and requirements. Local data protection rules (e.g. visibility of certain patients in certain departments) can be taken into account.

### Authorization of Components

Mutual access between IT components is defined in the respective configuration. To do so, the accessing system’s IP address and a password are recorded.

### User Authentication

User Authentication for the OSSE Registry for Rare Disease X is done via a username and password.

### Authentication of Components

Mutual access between IT components via the internet takes place only upon successful authentication. Authentication happens on the server side via server certificates and on the client side (depending on technical options) via IP address and username/password or via client certificates.

## IT Infrastructure Provisions

[*Apart from a description of the security measures that are part of the OSSE implementation, this section also includes suggestions/specifications concerning the organization and configuration of the IT environment. The statements have to be adjusted individually to the actual implementation.*]

### Security of Stored Data

All data collected in the central components of the OSSE Registry for Rare Disease X are stored on encrypted hard drive partitions. The corresponding key is located on a separate medium each per server (e.g. paper, USB stick). This medium is only required during mounting or booting and stored securely otherwise. Only the respective server’s administrator has access to “his/her” key medium. All servers are located in data centers furnished with access control via chip card or similarly secure tokens for authorized individuals.

### Security of Communication

Confidentiality of communication between components is ensured by the following measures:

* Communication between components generally occurs via encrypted connections (HTTPS). The keys and certificates used for this purpose must be generated in a way that corresponds to the current recognized requirements (e.g. key length). Current requirements can be found in the manuals for basic IT security of the German Federal Office for Information Security (<https://www.bsi.bund.de/DE/Themen/ITGrundschutz/ITGrundschutzKataloge/itgrundschutzkataloge_node.html>).
* Firewalls ensure that the servers running the central components can only be reached via those protocols and ports that are required for the communication with users or other components (usually HTTPS connections). Administrative access is restricted to the managing institution’s intranet.

### Logging

Access by researchers to components as well as access between components is logged. The record contains at least:

* The accessing person’s or component’s identity
* Access date and time
* Access content (transmitted data, in aggregated form if necessary) or information from which the content can be reconstructed (e.g. reference to a database entry or similar)

The record is stored together with the respective server’s payload for a period of one to six months. The recorded data must only be viewed for technical administration (particularly for troubleshooting) and in order to track abuse.

# Observing the Rights of Affected Individuals

## Information and Consent

[*Note on the scope of consent: If the consent form contains vague information e.g. on the storage period and use of the data, the patient information sheet must explicitly address and explain these issues.*]

The patient informed consent (see appendix for full text) provides the legal basis for data processing. With it, the patient particularly agrees that

* [***Version 1*** *and* ***version 2*** *(ID management exists)*] his/her identifying data are transferred to ID management and stored there,
* [***Version 3*** *(IDAT are manually assigned at the respective location)*] his/her identifying data are recorded manually in a list together with the internal PIDOSSE,
* Medical data and data on biomaterial samples are recorded in the OSSE registry according to the registry definition,
* Researchers of the OSSE Registry for Rare Disease X can analyze these data locally and search them via decentralized search,
* The patient’s medical data and data on biomaterial samples can be exported from the OSSE registry via a non-traceable export pseudonym and transferred to external researchers for those research purposes as specified in the informed consent.

The patient is informed of his right to information and revocation upon obtaining consent.

## Information on Stored Data

Patients recorded in the OSSE Registry for Rare Disease X have the right to receive information on the data stored in the registry. The information request has to be placed in writing and addressed to the treating hospital. The hospital requests a data export via the interface and receives an export pseudonym for the patient. The registry administrator carries out the export and produces a human-readable printout of the data, which he seals, marks with the export pseudonym and sends to the hospital in charge, where it can be handed over to the patient. If the data contain genetic results, it is mandatory that they be handed over in the context of a consultation by a treating doctor.

## Revocation, Deletion, Anonymization

Patients have the right to revoke their agreement on the processing of their data in the OSSE Registry for Rare Disease X. The revocation has to be placed in writing and addressed to the treating hospital, which passes it on to the registry board. Together with the revocation, the patient concerned can request the complete deletion of his/her data. If the latter request is missing, and if the data base allows for a factual anonymization, the data are anonymized. If “reasonable”[[12]](#footnote-13) anonymization is not possible due to low case numbers and specific characteristics, the data are deleted. This procedure excludes data that already form the basis of a published study.[[13]](#footnote-14) These data are then protected in a special way (e.g. archived separately) and access to them is denied.

[*The options and regulations concerning revocation, anonymization and deletion of data mentioned here (or further ones) have to be specified in the consent form and explained in the patient information sheet.*]

After reviewing the revocation, the registry commission decides on the request to deletion. In case of deletion, all data sets associated with the patient are deleted from the *Mainzelliste* and the OSSE registry. In case of anonymization, the data sets are deleted from the central patient list and the patient’s PSNOSSE is substituted with a random pseudonym. If data have been archived, this procedure is repeated for the archived data sets as well. The algorithm creating the pseudonyms ensures that the pseudonyms of a deleted/anonymized patient record are not used for new patient records.

The deletion or anonymization has to be carried out by the registry managers in charge promptly, no later than 14 workdays after the request was placed.[[14]](#footnote-15) The patient is informed of the completed deletion or anonymization in writing.

## Storage Period

The collected data will be stored in the OSSE registry as long as they can sensibly be used within the limits of the patient informed consent. In case the data can no longer be used as intended, the registry commission reviews whether there are legal grounds for a different use of the data, in anonymized form if applicable. If this review turns out negative, the data must be deleted.

# Appendix

## OSSE Registry Patient Informed Consent

[Consent template is inserted here]

## Data Sets

### Registry Definition

### Identifying Data

## Data and Source Systems for Data Import

1. TMF (Technology and Methods Platform) is a German platform and umbrella organization for networked medical research. Their guideline on data protection in medical research is available at <http://www.tmf-ev.de/Themen/Projekte/V039_01_WS_DS_Review.aspx>. This version of the guideline has received the following unanimous decision by the standing committee of German state commissioners for data protection on March 27 & 28, 2014 in Hamburg: *“The standing committee of federal and state commissioners for data protection recommends that medical research institutions and associations use the ‘Guideline on Data Protection in Medical Research. Generic Solutions of the TMF – Version 2‘, developed by TMF, as a basis for the concrete design of the data protection concepts.”* [↑](#footnote-ref-2)
2. For optimal use of the decentralized search, all participants (OSSE registries and OSSE bridgeheads) should employ a common identity management which can produce uniform export pseudonyms for their independent pseudonyms. It is crucial particularly in the area of rare diseases to not count identical patients multiple times in different registries. [↑](#footnote-ref-3)
3. This concept is based on the OSSE Data Protection Concept v1.2 by M. Muscholl, M. Lablans, A. Borg, F. Ückert and TOF Wagner. Translation by S. Buchberger. [↑](#footnote-ref-4)
4. *Mainzelliste* is the ID management software used in OSSE registries. It is a web-based pseudonymization service allowing for the creation of personal identifiers (PID). The term “*Mainzelliste”* will be used throughout this document to refer to an ID management concept implementing the Mainzelliste web interface. More information is available at [www.mainzelliste.de](http://www.mainzelliste.de). [↑](#footnote-ref-5)
5. This excludes deletions as per a patient’s request; see comments on the right of withdrawal in section 6.3. [↑](#footnote-ref-6)
6. With an OSSE bridgehead, registries not based on the OSSE registry software can be included in the decentralized search process. [↑](#footnote-ref-7)
7. ETL stands for “Extract-Transform-Load” and describes the technical and contentwise transfer of data from a source system to a target system. During the process, specific data adjustments (assignment to data fields, format changes, translation of values etc.) can be undertaken. [↑](#footnote-ref-8)
8. Pommerening, K, „Das Datenschutzkonzept der TMF für Biomaterialbanken“. ***it – Information Technology*** 49 (2007), 352–359. TMF is a German platform and umbrella organization for networked medical research. [↑](#footnote-ref-9)
9. The import of biomaterial data essentially follows the requirements of the TMF Data Protection Concept for Biorepositories. [↑](#footnote-ref-10)
10. I.e. individuals not affiliated with any of the OSSE Registry for Rare Disease X’s locations. [↑](#footnote-ref-11)
11. Usually, this should already have happened in the context of the individual’s work contract with the institution in charge. [↑](#footnote-ref-12)
12. Anonymization has to result in a minimum number of cases with the same characteristics so that patients cannot be identified based on their medical data. This is achieved e.g. through coarsening the characteristics into categories (e.g. age cohorts). Such coarsening only makes sense, though, if the original goal of the data processing can still be reached. [↑](#footnote-ref-13)
13. §20 German Federal Data Protection Act [↑](#footnote-ref-14)
14. The usually impracticable deletion or anonymization in data backups can be foregone if the backups can only be viewed by the system administrator in charge and old backups are deleted regularly. [↑](#footnote-ref-15)