

# OSSE – Open Source Registry System for Rare Diseases in the EU

Version: 2.0, March, 18, 2016

Marita Muscholl<sup>1</sup>, Dennis Kadioglu<sup>1</sup>, Martin Lablans<sup>2</sup>, Holger Storf<sup>3</sup>, Annette Pfalz<sup>4</sup>, Frank Ückert<sup>2</sup>, Thomas OF Wagner<sup>4</sup>

<sup>1</sup> Institut für Medizinische Biometrie, Epidemiologie und Informatik (IMBEI), Universitätsklinikum Mainz

<sup>2</sup> Medizinische Informatik in der Translationalen Onkologie, Deutsches Krebsforschungszentrum Heidelberg

<sup>3</sup> Medical Informatics Group (MIG), Universitätsklinikum Frankfurt

<sup>4</sup> Frankfurter Referenzzentrum für Seltene Erkrankungen, Universitätsklinikum Frankfurt

## 1. Summary

### 1.1 Background

In keeping with the European Council recommendations, Germany published a National Plan for Rare Diseases (RD) in August 2013, which includes 52 policy proposals to guide and structure actions in the context of rare diseases within the German health and social system. The project *OSSE (Open-Source-Registersystem für Seltene Erkrankungen in der EU / Open Source Registry System for Rare Diseases in the EU)*, funded by the German Federal Ministry of Health, provides a reusable software for RD registries.

### 1.2 Aim of the Project

*OSSE* provides patient organizations, physicians, scientists and other parties with open-source software for the creation of patient registries. As a result, the national registry landscape is empowered to comply with European principles regarding the establishment of minimum data sets, the compliance to data quality standards, etc. (summarized in the EUCERD Recommendation on RD registries). In addition, the necessary interoperability between different registries is supported from the start and allows to federate those registries on a national and international level (e.g. distributed searches designed to comply with data protection requirements and preserve data sovereignty).

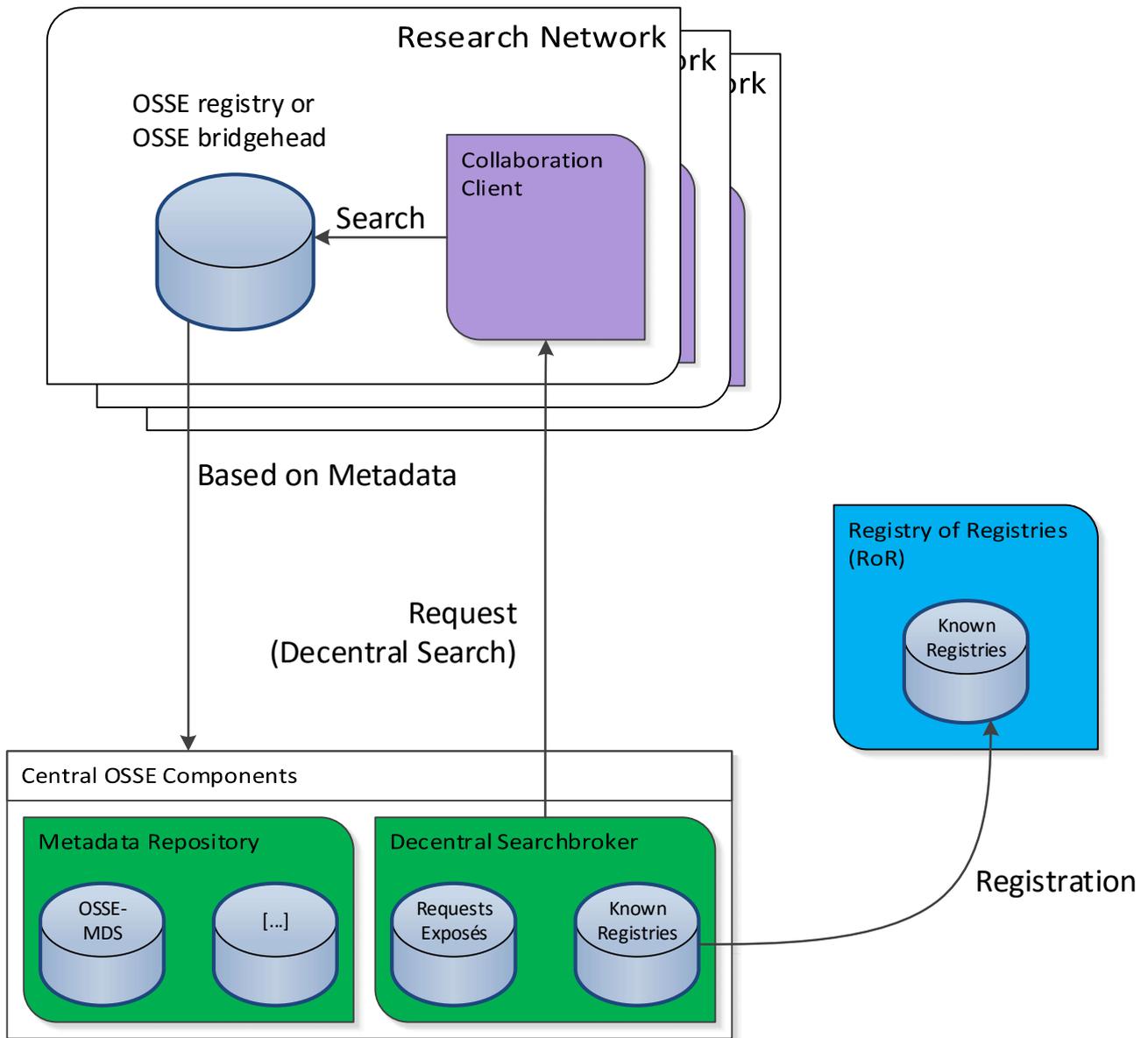
### 1.3 Concept

*OSSE* has defined and implemented a common data set (CDS) that should be held by all compliant registries<sup>1</sup>. Disease-specific additions can be collected and made available throughout the federated registries by defining them in a metadata repository (MDR). To meet the requirements of data protection, a template of required patient consent forms is provided along with a generic data protection concept taking into account the specificities of RD. A generic registry software was implemented and is distributed as an open source solution. It enables the handling of forms by providing an editor to read, modify and validate forms. Versioned persistency, secure authentication and an access management system is provided as well. As a proof of concept, the software is currently set up in two RD registries and will be employed during the first half of the year 2016.

---

<sup>1</sup> For the content of the CDS please refer to: [https://download.osse-register.de/MDS\\_3.0.pdf](https://download.osse-register.de/MDS_3.0.pdf)

## 2. Overall Concept



Architectural overview of the OSSE-Framework

### 2.1 The OSSE Registry Toolkit

The backbone of OSSE is a *registry toolkit* that enables scientists with a basic IT background to build a registry for a specific rare disease. A form editor allows defining forms for longitudinal and basic medical data and of the corresponding data schema. Each field (*including, inter alia, data type, ranges, measurement units and value sets<sup>2</sup>*) has to be defined within the metadata repository first.

### 2.2 The Metadata Repository

The integration of a metadata repository for rare diseases in the OSSE architecture eases the process of integrating data from different OSSE registries as all data set specifications for rare diseases used in the respective registries will be available through the MDR. If necessary, further items can be added while building national or regional disease-specific registries and will also be available for all OSSE registries. The metadata repository component also

<sup>2</sup> according to ISO/IEC 11179 Metadata Registry Standard

allows for the retrieval of metadata items from other MDRs. Apart from self-defined value sets, the MDR also provides access to standardized codes such as ICD-10-GM or Orphacode (added later).

## 2.3 The OSSE Registry

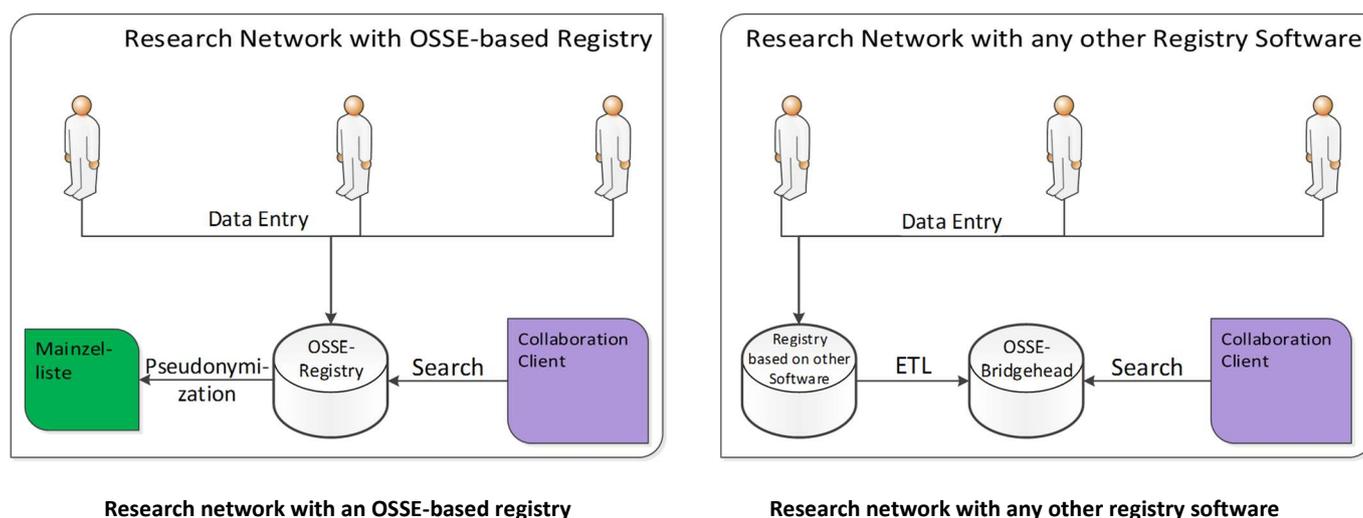
The OSSE registry provides role-based access control, plausibility checks regarding the validity of entered values, data versioning and a first workflow support through different statuses and allowed status transitions per role. An interface for a configurable data import and data export is available as well. For better discoverability, each registry should register with a *registry of registries (RoR)*.

## 2.4 The “Distributed Search” Principle

It is our strong belief that raw patient data should not leave the local registry even with the presence of patient consent, because there is low acceptance among patients and data owning scientists of big data collections and their controllability particularly outside of the influence of national data protection rules. Instead, we developed a concept for a distributed search which makes data from (national and regional) rare disease registries available, while respecting data ownership and privacy aspects. The *search broker* allows specified search queries based on the existing MDR items. An exposé describing the research question and supplying contact information of the inquiring partners completes the request. The local request interface of each OSSE registry, called the “*Collaboration Client*”, runs the received inquiries. If there is a non-empty result set, it is presented to the person in charge of each matching registry together with the exposé and the inquirer’s contact information. Finally, the data owner decides if and what to reply.

## 2.5 Integration of Registries with the OSSE Bridgehead

In order to enable the integration of registries built with a different software, a so-called “*OSSE bridgehead*” can be installed. The OSSE bridgehead consists of the OSSE core components including the query interface to the Collaboration Client. Data has to be entered from the registry by a periodically running ETL<sup>3</sup> process that can optionally access the local ID management, if the creation or modification of patient pseudonyms is necessary. During the transformation step, each attribute has to be mapped to a data element defined in the MDR.



<sup>3</sup> ETL means “Extraction, Transformation, Load”, the basic steps feeding the bridgehead.

## 2.6 Pseudonymization

Data protection regulations in Germany (and also in Europe) require pseudonymization as a basic prerequisite for the registration and storage of patient data in research networks. The part of the data that allows for the identification of patients (e. g. family name, first name, date of birth, name at birth, etc.) has to be replaced by a pseudonym and stored together with the pseudonym in a separate patient list deployed on a different server which is controlled by a trusted third party. Every time data for a patient is entered, the identifying data has to be sent to the identification management where special record linkage algorithms verify the existence of an identical or very similar data set. The corresponding ID/pseudonym will be returned to the requesting system. Research networks with different data stores, e. g. a registry, a research database and a biomaterial database, have to assign different pseudonyms for each system (according to German data protection laws).

In Germany, TMF e.V., a technology and methods platform and umbrella organization for networked medical research, developed a comprehensive data protection handbook for medical research networks [1], which describes, among other things, requirements for the ID management and pseudonymization in medical research networks. In compliance with these requirements, the OSSE architecture includes a local ID management/pseudonymization service for each national or regional OSSE registry. The ID management software that we use in OSSE, called “Mainzelliste” [2], has been released as open source by the Medical Center of Johannes Gutenberg University Mainz (Division of Medical Informatics) and is already used in various other projects.

## 3. References

[1] K. Pommerening, U. Sax, T. Müller, R. Speer, T. Ganslandt, J. Drepper, S. Semler, Integrating eHealth and medical research: The TMF data protection scheme. In: B. Blobel et. al (Eds): *eHealth: Combining Health Telematics, Telemedicine, Biomedical Engineering and Bioinformatics to the Edge*. Aka, Berlin 2008, 5–10

[2] Die Mainzelliste als OpenSource. [ <http://www.unimedizin-mainz.de/imbei/informatik/opensource/mainzelliste.html> ], last visited February 2016

## 4. Contact

Dr. Holger Storf  
Universitätsklinikum Frankfurt  
Medical Informatics Group (MIG)  
Haus 33C 2 OG R219  
Theodor-Stern-Kai 7  
D-60590 Frankfurt  
Tel: +49 69 6301-84438  
Mail: [storf@med.uni-frankfurt.de](mailto:storf@med.uni-frankfurt.de)