Lucerne University of
Applied Sciences and Arts

HOCHSCHULE
LUZERN

**Informatik**
FH Zentralschweiz

# Optimus Study Cycle 3 – IT Security Concept

**Jana Koehler, Stefan Schnürle, Roland Portmann**

## Abstract

The main focus of this IT security concept is to strictly secure the upload of and access to the sensitive data collected for the Optimus Study Cycle 3. The institutions participating in the study will upload their case data. Case data can occur in non-anonymized form, including personal and very sensitive information if an institution can only provide the data in the way they are originally stored. Some case data might be anonymized depending on the software export capabilities of the tools used by an institution. This security concept considers the most challenging scenario of original case data and provides solutions to protect these data effectively:

- Access limited to only pre-registered users requiring 2-factor-authentication.
- Immediate removal of case data from the internet infrastructure and storage in a highly secured and encrypted backend system (deletion after end of the study),
- Automated extraction of anonymized study data, which removes all sensitive and personal information.
- Access of all users limited to study data based on a need-to-know and role-based access control.

# Technical Report 2/2017

DOI: 10.5281/zenodo.1288550

## The Technical Report Series

Technical Reports in this series publish research results and working papers from the School of Information Technology at Lucerne University of Applied Sciences and Arts covering a wide range of topics.

## Contact

**Hochschule Luzern – Lucerne University of Applied Sciences and Arts**
**Informatik – School of Information Technology**

Suurstoffi 41b
Ch-6330 Rotkreuz
Switzerland
www.hslu.ch/informatik

## Impressum

Edited by the School of Information Technology at Lucerne University of Applied Sciences and Arts.

UNIL | Université de Lausanne
Observatoire maltraitance envers les enfants

Lucerne University of Applied Sciences and Arts
HOCHSCHULE LUZERN

# Optimus Study Cycle 3 – IT Security Concept

28.01.2016

*Jana Koehler*[a], Dr. habil., *Stefan Schnürle*[a], BSc, & *Roland Portmann*[a], dipl. Ing. ETH

[a] Lucerne University of Applied Sciences and Arts – School of Engineering and Architecture

The main focus of this IT security concept is to strictly secure the upload of and access to the sensitive data collected for the project. The institutions participating in the study will upload their **case data**. Case data can occur in non-anonymized form, including personal and very sensitive information if an institution can only provide the data in the way they are originally stored. Some case data might be anonymized depending on the software export capabilities of the tools used by an institution. This security concept considers the most challenging scenario of original case data and provides solutions to protect these data effectively:

- Access limited to only pre-registered users requiring 2-factor-authentication.
- Immediate removal of case data from the internet infrastructure and storage in a highly secured and encrypted backend system (deletion after end of the study),
- Automated extraction of anonymized **study data**, which removes all sensitive and personal information.
- Access of all users limited to study data based on a need-to-know and role-based access control.

## Contents

# 1. Overview over the Infrastructure

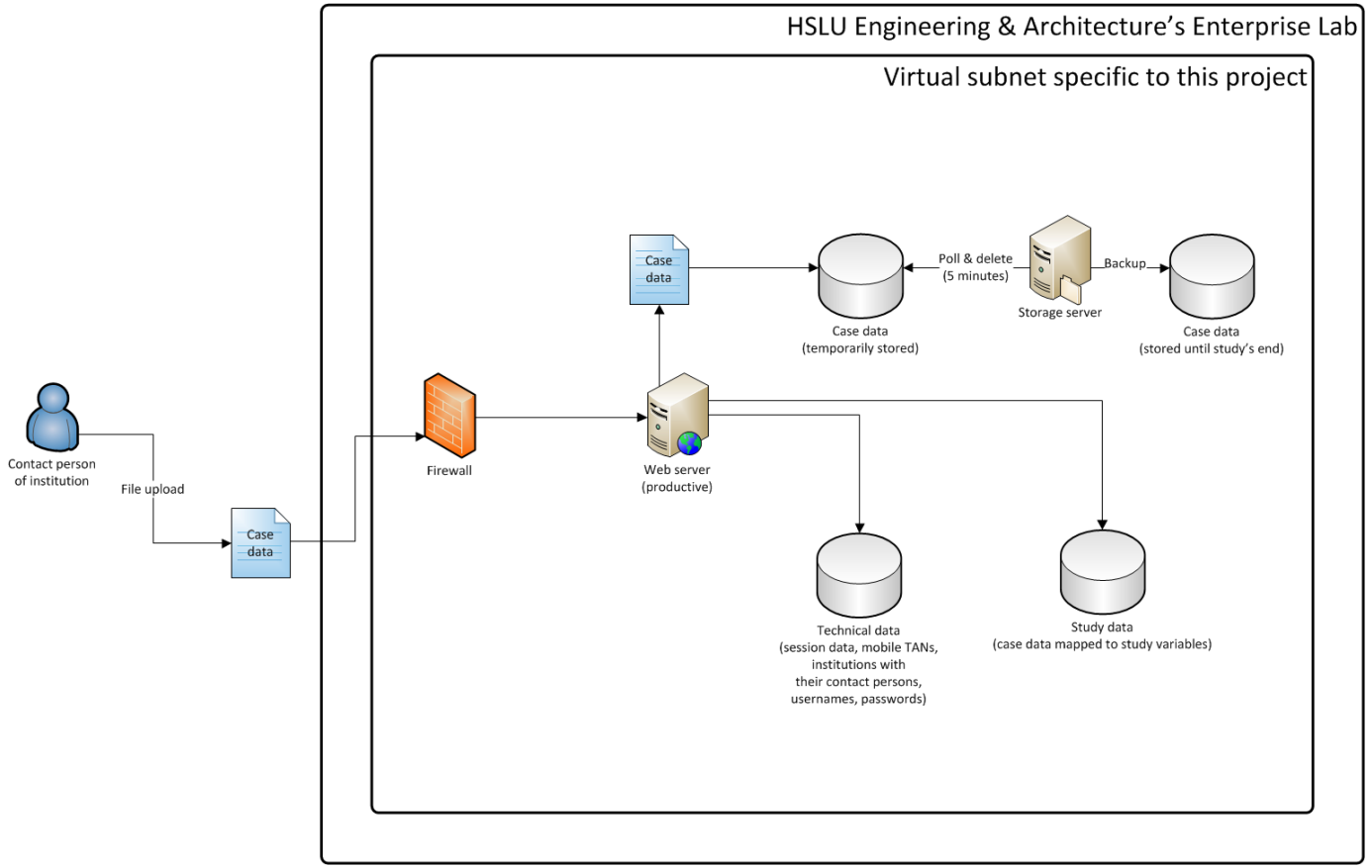


Figure 1: Service „Upload case data“

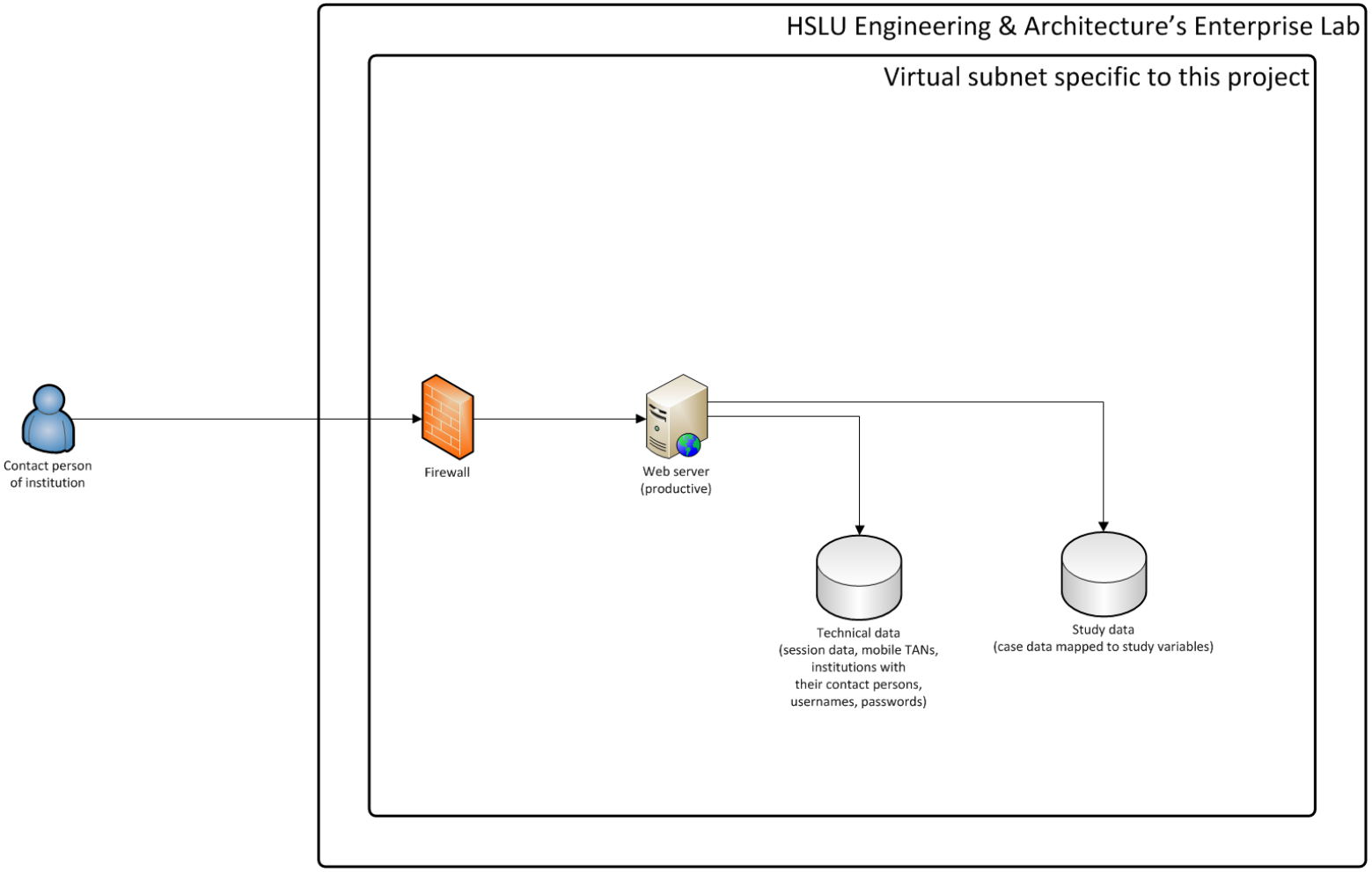


**Figure 2: Service „Complete study data“**

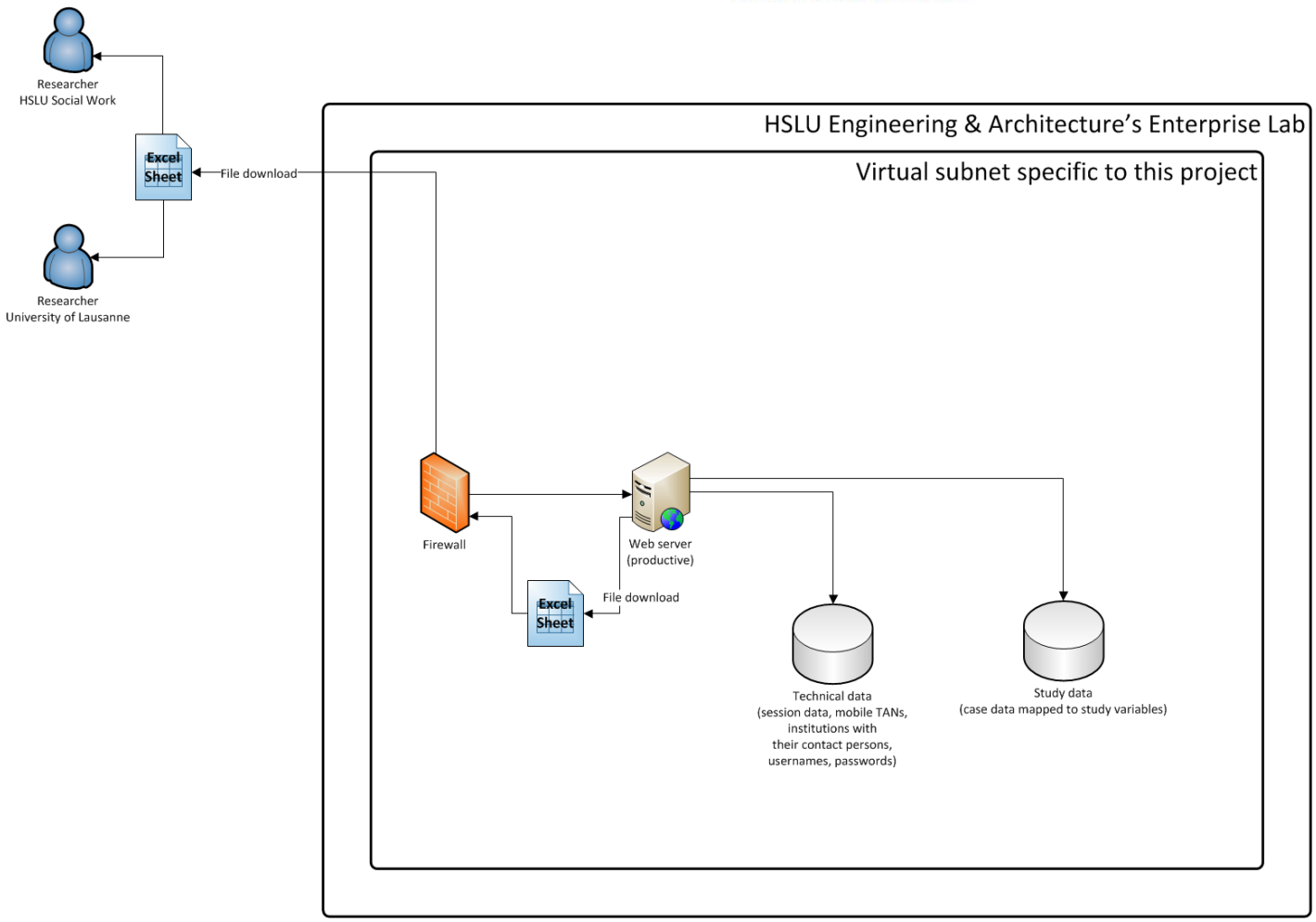


**Figure 3: Service „Analyse study data“**

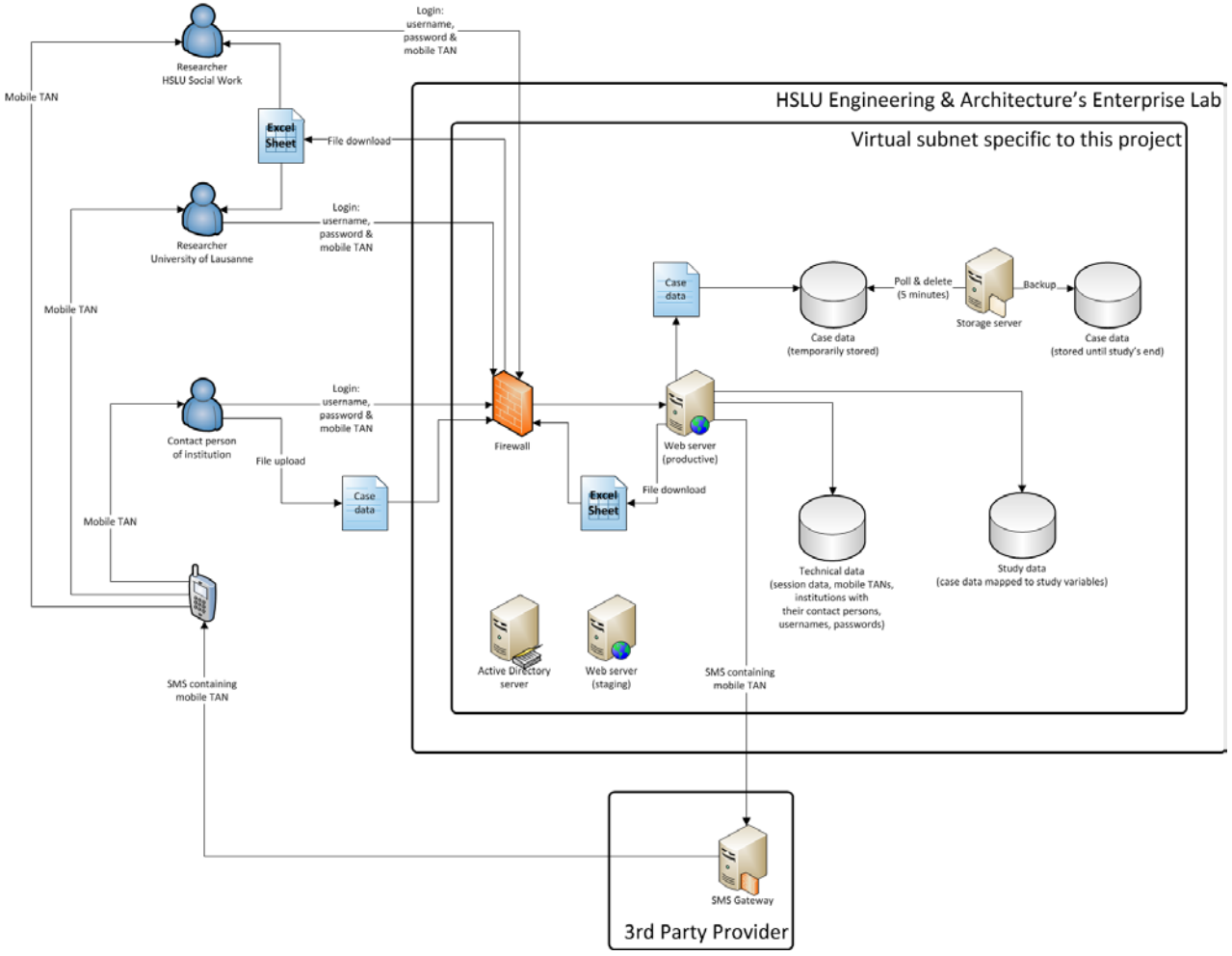


**Figure 4: Complete infrastructure**

# 2. Interaction of Users with the Infrastructure

Different *institutions* (e.g., child protection groups, social and legal services, hospitals, etc.) participate in the study and will provide data. The participating institutions are chosen by the Lucerne University of Applied Sciences and Arts – School of Social Work by means of a stratified sample. Each institution appoints one contact person, who is responsible for the use cases "upload case data" (see section 2.1) and "complete study data" (see section 2.2).

## 2.1. Service „Upload Case Data"

Prerequisites:

- A contact person is authenticated and logged in to the system (see chapter 3). See Figure 1.
- The contact person uploads the case data of the institution once.

The contact person is able to upload a file via a web frontend containing the case data exported from the institution's software tool. The format of the data (e.g. Excel, CSV, XML file, other) is determined beforehand for each institution depending on their tool's capabilities.

Upon successful upload, the case data is immediately processed to extract the study data using a mapping specific to each institution. The study data is stored in the database of the web server and does not contain any personal data potentially contained in the case data. The contact person is presented the extracted study data via a web form in order to enable her/him to verify the correctness of the imported data. If necessary, she/he can edit resp. complete the study data, see section 2.2. No access to the case data is possible via the infrastructure.

In intervals of 5 minutes, a separate process running on the storage server in the encrypted backend of the infrastructure connects to the database on the web server. This process copies the case data to the storage server's database and irrevocably deletes the case data from the web server's database.

We assume that an upload only happens once for each institution. If an institution repeatedly uploads case data, all their previously extracted study data is overwritten and replaced by the study data extracted from the most recent upload. No versioning of the data or data history is created.

## 2.2. Service „Complete Study Data"

Prerequisites:

- A contact person is authenticated and logged in to the system (see chapter 3). See Figure 2
- Upload and mapping of the case data of the institution was successful.

The contact person can access a web form via the same web frontend by which she/he uploaded the case data (see section 2.1) and edit resp. complete the extracted data. Since the study data is stripped of all personal data which may have been present in the uploaded case data, the contact person has to identify each case by its characteristics, e.g. by the date the case was first reported, or variables such as age or gender.

As soon as the contact person saves the data, the corresponding record in the study data is updated. No versioning of the data or data history is created; the study data in the database contains the most recent edit via the web form (see section 2.1).

### 2.3. Service "Analyse Study Data"

Prerequisites:

- A researcher is authenticated and logged in to the system (see chapter 3). See Figure 3

A restricted and pre-registered number of researchers, the members of the study team for the Optimus Study Cycle 3 from the Lucerne University of Applied Sciences and Arts – School of Social Work and the University of Lausanne, are granted access to a web frontend that provides a link to download an Excel sheet generated automatically from the database containing the study data.

## 3. Authentication and Authorisation

The services provided by the infrastructure are secured in the following way:

Authentication of all users is secured via two-factor authentication. All users – contact persons as well as researchers – log in by providing their username and password via web frontend secured by a HTTPS connection. After successfully validating these credentials, a transaction number (TAN) is sent via SMS to the mobile phone assigned to this user. This individual TAN is randomly generated by the web server and sent via a 3rd party provider. After the user entered the correct TAN in the login form, the two-factor authentication is completed, the user is logged in to the system and is able to use the services for which she/he is authorised. After 3 failed login attempts, the user's account is locked for 15 minutes in order to impede brute force attacks.

The username and the initial password are sent to each user by surface mail letter. After initial log in to the web frontend, the user is asked to reset the personal password. Password complexity is enforced by requiring a minimum length of 8 characters, at least 1 lowercase character, at least 1 uppercase character, at least 1 special character, and at least 1 number. Only the hash created by the algorithm specified in RFC 2898[1] is stored in the user database.

There is no functionality available on the web server to register new users. User registration is handled with separate software by technical personnel of Lucerne University of Applied Sciences and Arts - School of Engineering and Architecture only.

The authorisation of the users is controlled by a role-based access control (RBAC); the institutes' contact persons are allowed to use services 1 and 2, whereas the researchers from the study team are only allowed to use service 3.

[1] see http://www.ietf.org/rfc/rfc2898.txt, https://msdn.microsoft.com/en-us/library/system.security.cryptography.rfc2898derivebytes%28v=vs.110%29.aspx

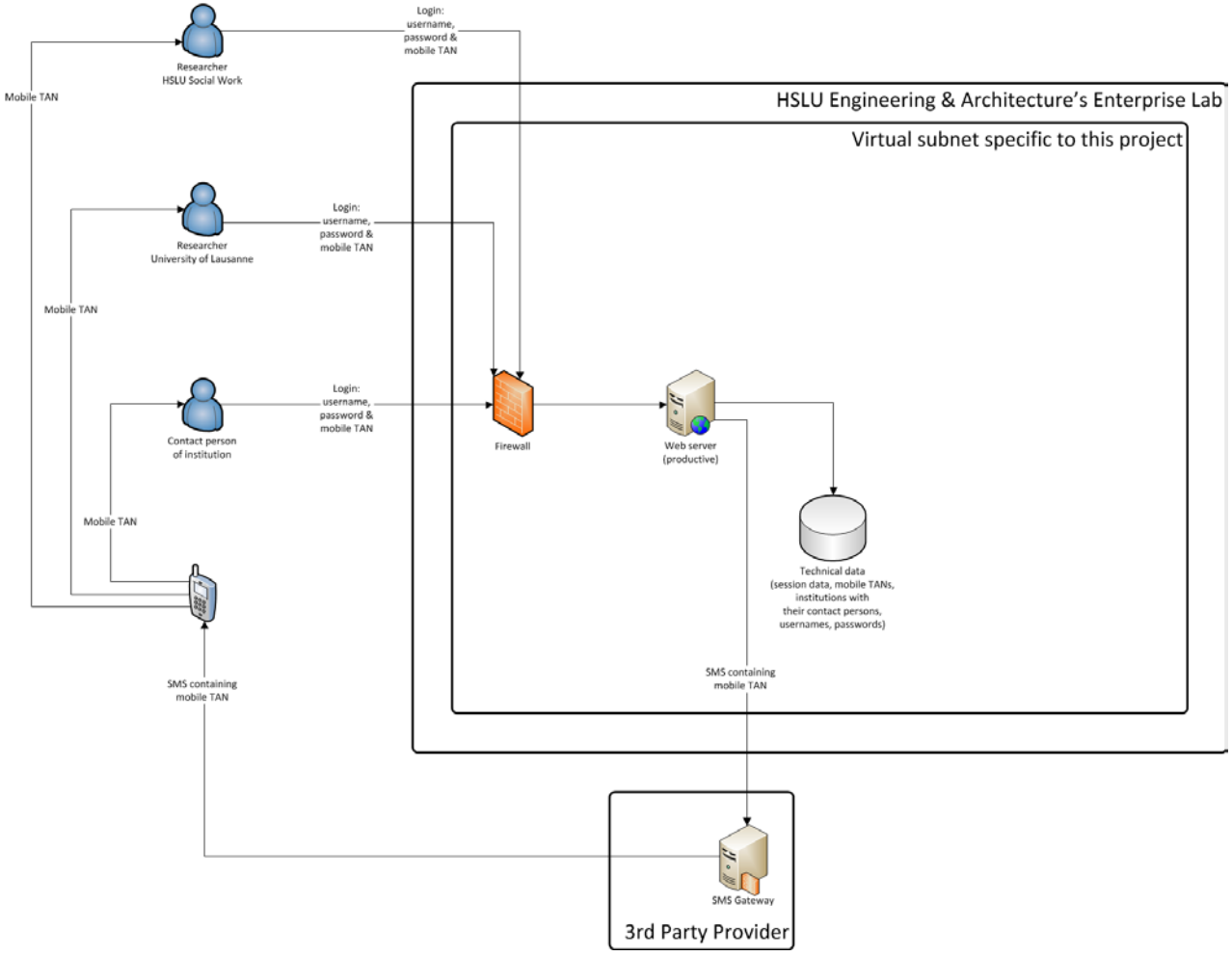


**Figure 5: Authentication and authorisation**

# 4. Secure Hosting of the Project Infrastructure

The described infrastructure is hosted in a subnet, which is an isolated, virtual demilitarized zone (DMZ) within the Enterprise Lab of the Lucerne University of Applied Sciences and Arts – School of Engineering and Architecture. Only services 1 to 3 are provided to the internet. These services are limited to access the web server only, which is secured by an encrypted connection via Transport Layer Security (TLS) at TCP port 443 (see Figure 6). Direct access to other services like, e.g., direct access to the database via an SQL client or to the storage server is thus impossible.

The technical data stored in the corresponding database comprises data of pre-registered users, information on the mapping to be used for data processing, and session data recording activities on the web server.

The team of Lucerne University of Applied Sciences and Arts – School of Engineering and Architecture will designate 1-2 technical administrators, who are responsible for the operation of the infrastructure and for providing support to the institution's contact persons. They have to sign a non-disclosure agreement. These administrators are authenticated against the Active Directory domain within the subnet which has no trust relations established to other domains (see chapter 7).

All servers within the virtual subnet are virtual machines. Their virtual hard disks (which are encrypted by Windows' BitLocker technology) are regularly backed up to the data store provided by IT Services of Lucerne University of Applied Sciences and Arts. The backed up hard disks are never transferred to entities outside of the university.

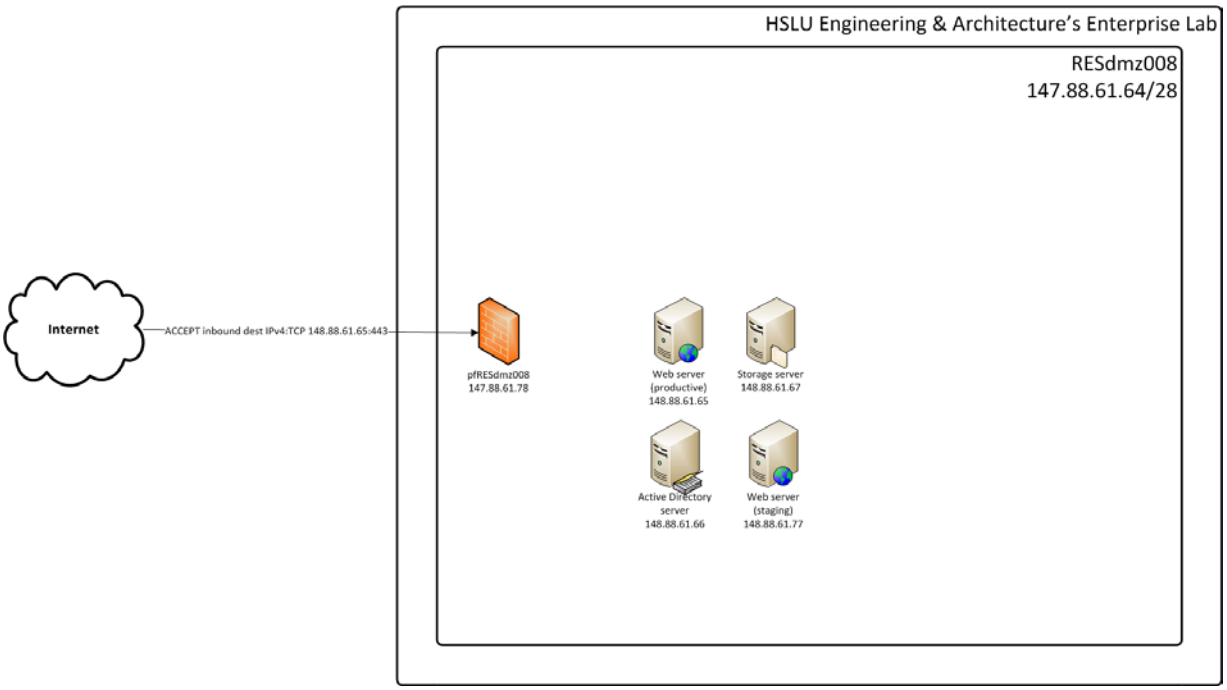


**Figure 6: Technical details of the virtual subnet specifically set up for this project**

# 5. Detailed Information on the Server Infrastructure

Each server is stored on a virtual hard disk, which is encrypted using Windows' BitLocker technology. The BitLocker recovery key is stored within a secured document space in the Enterprise Lab of the Lucerne University of Applied Sciences and Arts – School of Engineering and Architecture.

## 5.1. Web Server (productive)

The web server is accessible via HTTPS only. Its services are available only after successful two-factor authentication (see chapter 3). Besides the login form, the web server offers no public services to the internet.

The uploaded files containing case data of the institutions are temporarily stored in a database on the web server. Every 5 minutes, a separate process running on the storage server connects to this database on the web server, extracts the case data, stores it in the storage server's database and deletes the case data from the web server. Hence, the time is minimized when potentially sensitive personal data is stored on the web server. The web server is undergoing a penetration test before going productive.

## 5.2. Storage Server

The storage server is used to back up the uploaded case data. It is not accessible via the Internet.

## 5.3. Active Directory Server

The Active Directory server is used to authenticate and authorise personnel with administrative access to the servers within the virtual subnet. This access is necessary for the operation of the infrastructure and to provide support for institution's contact persons. Every person granted this access has signed a non-disclosure agreement.

## 5.4. Web Server (Staging)

The web server for staging for testing is set up the same way as the productive server and provides a means to conduct tests of new versions of the application before they are deployed to the productive web server. On this staging web server, no case data potentially containing sensitive, personal information is stored.

# 6. Logging

All actions of users conducted via the web application are logged in the technical database.