Project title	Artificial intelligence and the personalized prevention and management of chronic conditions		
Project acronym	WARIFA		
Project number	101017385		
Call	Digital transformation in Health and Care	**Call ID**	H2020-SC1-DTH-2020-1
Topic	Personalised early risk prediction, prevention and intervention based on Artificial Intelligence and Big Data technologies	**Topic ID**	SC1-DTH-02-2020
Funding scheme	Research and Innovation Action		
Project start date	01/01/2021	**Duration**	48 months

D1.11 – DATA SECURITY AND ANONYMISATION MEASURES

Due date	30.06.2021	**Delivery date**	30.06.2021
Work package	WP1 Project Management		
Responsible Author(s)	Conceição Granja Bartnæs		
Contributor(s)	All partners		
Version	1.0		

DISSEMINATION LEVEL

Please select only one option according to the GA			
☒	PU: Public	☐	PP: Restricted to other program participants
☐	RE: Restricted to a group specified by the consortium	☐	CO: Confidential, only for members of the consortium

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017385

VERSION AND AMENDMENTS HISTORY

Version	Date (MM/DD/YYYY)	Created/Amended by	Changes
0.0	22.04.2021	Merethe Drivdal	Creation of document
0.1	05.05.2021	Merethe Drivdal	Update on all chapters
0.2	08.05.2021	Conceição Bartnæs	Update
0.3	27.05.2021	Eva Henriksen	DPO Input
0.4	26.06.2001	Conceição Bartnæs	Update

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017385

TABLE OF CONTENTS

1 INTRODUCTION 5
2 DATA SECURITY AND ANONYMISATION MEASURES 5
2.1 DATA STORAGE 5
2.2 ACCESS RIGHTS AND PROCEDURES 6
2.3 PRIVACY PROTECTION 7
2.4 FAIR DATA MANAGEMENT 7
2.5 DOCUMENT MANAGEMENT 7
2.6 DATA ARCHIVING AND PRESERVATION 7

LIST OF FIGURES

Figure 1 TSD registration form for foreign project members. 6

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017385

List of Abbreviations

Abbreviation	Meaning
AI	Artificial Intelligence
DMP	Data Management Plan
TRL	Technology Readiness Levels
TSD	Services for Sensitive Data (Tjenester for sensitive data)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017385

1 INTRODUCTION

The WARIFA project aims to facilitate personalised early risk prediction, prevention and intervention based on Artificial Intelligence (AI) and Big Data technologies. We want to explore how AI-based mobile applications may be used as a tool for individual lifestyle changes. This includes the use of mobile applications to analyse and estimate individual risk, correlate it with the community risk profile, provide evidence-based and personalized advice together with prompts for preventive lifestyle changes. The aim is to empower citizens to self-monitor the implementation of risk-reducing lifestyle changes. WARIFA will develop an AI-based system with the aim to help prevent chronic conditions for all citizens.

By combining ubiquitous data from the user`s environment with user-generated data, AI algorithms can process the most relevant data in the appropriate context and then provide the tools for personalized advice resulting in more specific preventive interventions. AI and big data technologies have the potential to address these challenges by analysing risk levels and providing citizens with tailor-made advice according to the individual risk level.

To achieve this objective it is necessary to combine ubiquitous data and personal user-generated data, and to combine interdisciplinary efforts from clinical, technical, and sociology background, in order for the WARIFA prototype to reach TRL 6-7 by the end of the project period. The development of the WARIFA the system will be iterative with respect to design/development /testing/feedback-adjustments. Human subjects will be involved in several steps of the project.

The Participant Identification and Recruitment Plan provides the ethical guidelines and procedures that will govern the WARIFA project activities. This deliverable describes the procedures and tools of the project to ensure good research practices when involving end-users and stakeholders, and it serves as a reference to which all the WARIFA project members have to comply in their participation in project activities.

The Participant Identification and Recruitment Plan shall be updated, as necessary, during the duration of the project, and is kept available for all WARIFA project members on the chosen platform for project interaction, Microsoft Teams.

2 DATA SECURITY AND ANONYMISATION MEASURES

2.1 DATA STORAGE

All sensitive data collected during the course of the WARIFA project will be stored at TSD (from the Norwegian, Tjenester for Sensitive Data). TSD offers storage and processing of data in a secure environment. The solution is run on dedicated computers in a separate location in USITs machine room to which only USITs operational personnel has access. To achieve complete separation of project environments running on the same hardware, we use RHEV KVM as a hypervisor. This means that a physical computer can be divided into several separate virtual computers which for all intents and purposes are working independently. A two-step authentication is needed to gain access to the system. Inside the system, every project has its own VLAN and its own virtual file system. This means that projects cannot find any information about any other project on the system. A risk analysis of TSD is provided in Annex 1.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017385

2.2 ACCESS RIGHTS AND PROCEDURES

Only the members of the WARIFA consortium will have access to the project data. Data transfers to and from TSD are handled by a special purpose file staging service and the project coordinator controls access rights for all project members. By default all project members are able to transfer data in, but only the project coordinator can transfer data out.

Project members resident in Norway will be able to login to TSD services using their BankID for identity verification. Electronic identification using BankID meets the official requirements that apply to identity verification and binding electronic signature, providing access to the TSD services at the highest level of security, level 4. Foreign project members must be registered by the project coordinator to be granted access to the WARIFA project space at TSD (Figure 1). All changes in access rights require a written approval from the project coordinator.

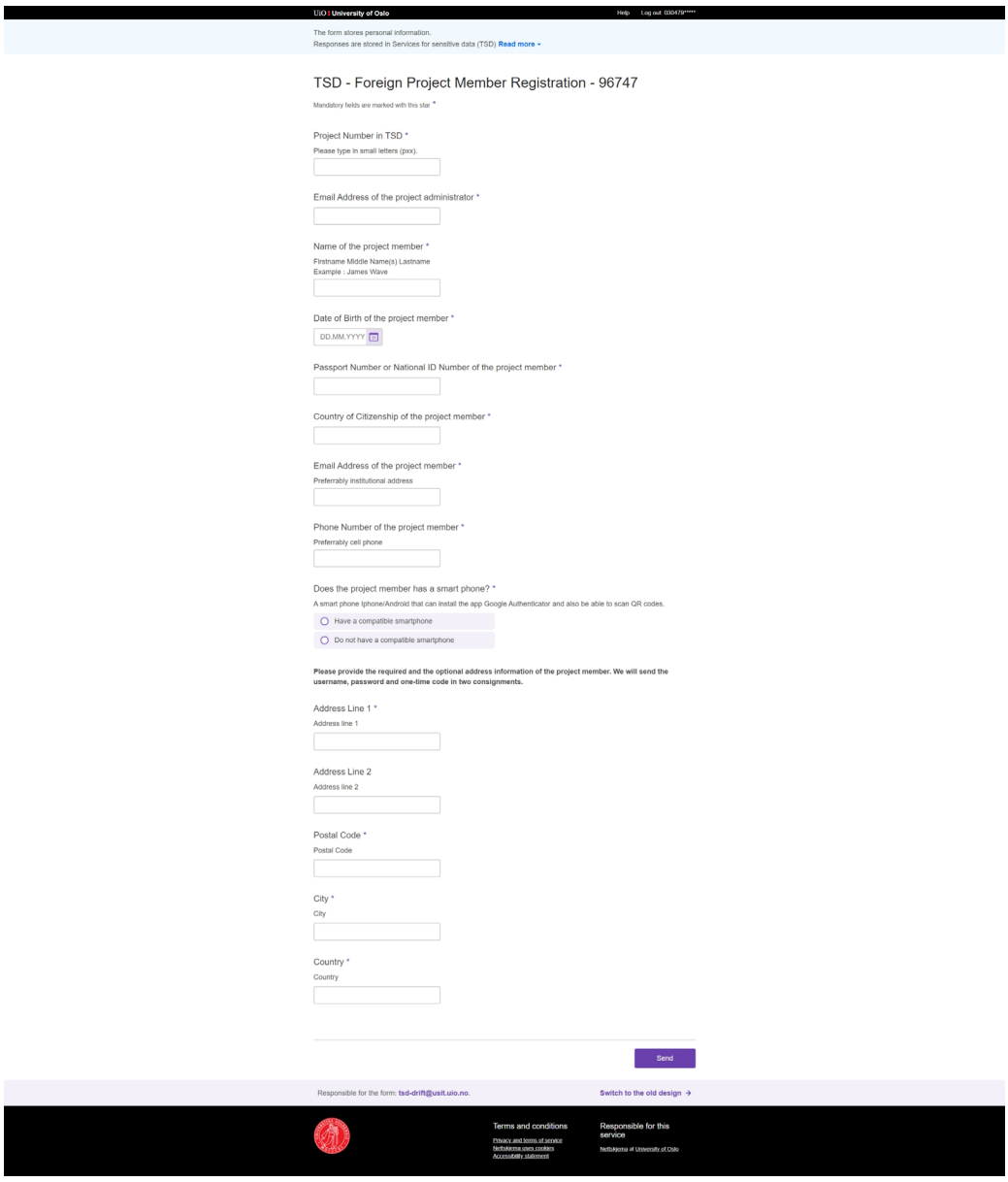

Figure 1 TSD registration form for foreign project members.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017385

2.3 PRIVACY PROTECTION

To minimise storage of personal data, WARIFA will implement the following Pseudo-anonymization procedures. Participants will be identified by a unique identifier number. This identifier number will be used to denote individual participants within all data files. A key-file containing identity information will be created which links each unique identifier number with the participant's initials and date of birth. This key-file will be password-protected, encrypted and saved on the secure institutional server, separate from the rest of the data. Table 1 describes the anonymization level of the datasets collected and generated during the WARIFA project activities.

Table 1 Description of the level of anonymization of the datasets collectged and generated in WARIFA.

Dataset Name	Level of Anonymization
WP2 Mapping	Anonymous
WP3Sensor data	Pseudonymous Data
WP3 Ubiquitous data	Pseudonymous Data
WP3 Registry data	Pseudonymous Data
WP7 Bayesian belief network	Pseudonymous Data
WP7 Risk scores	Pseudonymous Data
WP8 Surveys and interviews	Anonymous

2.4 FAIR DATA MANAGEMENT

The FAIR Data Principles (Findable, Accessible, Interoperable, Reusable) have received worldwide recognition as a useful framework for thinking about sharing data in a way that will enable maximum use and reuse. The adherence to the principles: (1) supports knowledge discovery, innovation and knowledge integration, (2) promotes sharing and reuse of data across disciplines, (3) supports new discoveries through the harvest and analysis of multiple datasets. WARIFA will follow the Guidelines on FAIR Data Management in Horizon 2020 in the best possible way.

2.5 DOCUMENT MANAGEMENT

Non-sensitive files and documents produced by project members can be shared among the WARIFA consortium in a repository. The repository (locally hosted SharePoint server) is provided by the project coordinator. The SharePoint servers are a part of the Microsoft 365 suite and are hosted in Europe. The repository is accessible by invitation from the project coordinator, and access will only be granted to registered members within the consortium. Sensitive data is not to be shared on the Teams platform. Sensitive data will be stored at TSD as described on section 2.1.

2.6 DATA ARCHIVING AND PRESERVATION

To comply with the rules of good scientific conduct and the related storage of sensitive data, consent forms, and project files and documents within WARIFA will be stored at TSD for 5 years after the end of the project.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017385

Annex 1 Risk Analysis of TSD

Risk Analysis of TSD - Tjenester for Sensitive Data

Version 4.0
2018-03-08

Espen Grøndahl
IT-Security Officer
UiO

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017385

1. INTRODUCTION 3
2. BACKGROUND 3
3. LIMITATIONS 3
4. SHORT DESCRIPTION OF THE SOLUTION 3
5. SECURITY REQUIREMENTS 5
6. SECURITY ASSESSMENT 5
6.1 METHOD 5
6.2 MEMBERS 5
7. ACCEPTANCE CRITERIA 6
7.1 EVALUATION SCALES 6
7.2 RISK ELEMENTS BETWEEN 1 AND 3: 7
7.3 RISK ELEMENTS BETWEEN 4 AND 8: 7
7.4 RISK ELEMENTS BETWEEN 9 AND 16: 7
8. DESCRIPTION AND ASSESSMENT OF THE RISK ELEMENTS 7
RISK ELEMENT 5 7
RISK ELEMENT 7 8
RISK ELEMENT 14 8
RISK ELEMENT 15 8
RISK ELEMENT 16 8
RISK ELEMENT 19 8
RISK ELEMENT 20 8
9. FINAL ASSESSMENT AND FOLLOW UP 9
10. CONCLUSION 9
11. ATTACHMENTS 9

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017385

1. Introduction

TSD - Services for Sensitive Data (Tjenester for Sensitive Data in Norwegian) was developed by USIT at the University in the period 2011-2014. It is based on a pilot TSD 1.0 which was made 2008-2011. The system offers a secure environment that meets all legal requirements regarding privacy and protection of sensitive data. The service is meant to host primarily sensitive research data within the health sector, but also other data that require additional protection.

Since it was launched in May 2014 the service is in continuous development to meet the evolving needs of the targeted user communities. This document is the fourth version of the risk assessment and is based on TSD as of February 2018. The previous version of the risk assessment was published in February 2017. This most significant change to the core infrastructure is the addition of the TSD HTTP API for data transport. We have also made changes to administrative procedures which will be noted throughout the text.

2. Background

TSD is in full production from May 2014. This risk assessment is the fourth general assessment of TSD and will be the basis for the regulatory risk which each project manager will have to make before they use the system for storage and processing of sensitive personal data.

3. Limitations

This risk assessment addresses TSD and its major components, as well as surrounding infrastructure that might have direct or indirect impact on the solution. The document does not provide details of the single components, unless it is important to describe the overall risk picture. The technical description of each component is provided in the attached Whitepaper.

This document does not cover the risks associated with special solution implemented to address the needs of single projects, but is based on a general usage patterns. Projects that use the system must supplement with their own judgments which are specific to their projects.

4. Short Description of the solution

A detailed description of the solution can be found in the attached whitepaper and on the service web-pages. Some details cannot be disclosed and are provided only on request.

TSD has physically and logically closed network with strong access control, and operations are performed using automation tools where possible. The infrastructure

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017385

consists of virtualized linux and windows environments dedicated to each project, running on a dedicated project VLAN, connected to the central storage facilities. A number of dedicated VLANs are occupied by administrative virtual machines. A cluster of three Active Directory (AD) instances, one physical and two virtual, running on one of the administrative VLANs, is used for user and group administration, and the distribution of host policies. The project machines can only communicate with the resources on their own project VLAN and have no connection with other VLANs or outside networks. Projects do have some openings to the administrative VLANs to enable usage of shared network services and resources. The only network traffic allowed between the administrative machines and the "world" is for the actual user access, logs and license info.

Users access TSD by establishing an encrypted communication to dedicated gateway machines. Login protocols are based on PCoIP (to log on windows machine) or Thinlinc (linux machines). Login requires Two-Factor Authentication using username, password and a one-time code. A person having access to two or more different projects will have a dedicated username per project in order to avoid leakage of data through projects. Functionalities that can enable uncontrolled transfer of data (cut-and-paste functionalities, the mapping of local drives, USB forwarding) are disabled.

All network traffic into and out of TSD is controlled by the firewall. This includes login and data transfer. The latter can be done in one of two ways: 1) using the filelock and/or 2) using the TSD HTTP API.

The Filelock consists of two machines, one physical machine, accessible by the user through a SFTP channel upon two factor-login, and one virtual machine, on the inside of the TSD network that can mount projects areas. The two machines are synchronized, so that a file placed from the outside on the physical machine is automatically copied to the inside and vice versa, in case of export. The Filelock mechanism allows strong control of user import/export privileges and logging of files name and eventually checksum data. The solution is designed to create several barriers towards attack so to minimize the impact of a potential security breach.

The TSD HTTP API also allows data import and export. It supports three authentication methods: basic authentication, TSD two factor authentication and BankID authentication. All API clients are verified by TSD staff and are given access to specific projects on a per-need basis. Basic authentication only allows data import and is only allowed from API clients running on machines with specific IP addresses. The API is integrated with TSD's internal Identity Provider and authentication and authorization system. Two factor authentication using TSD credentials and BankID credentials allow for both data import and export, in principle. The same authorization system applies whereby a project administrator has to grant export rights to a project member. In addition to user authentication, applications have to authenticate themselves towards the API, using a revocable time-limited API key. This allows TSD to revoke access to any application at any time, should it be necessary.

All the users and administrative personnel must log onto the TSD via two factor authentication. Technicians and sys-admins have a user credential so that every operation can be traced and eventually connected to the responsible person.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017385

5. Security requirements

The degree of security is evaluated often along three axes. Confidentiality - ensuring that no one is able to access the data other than those who have legitimate needs. Integrity - ensuring that the data or the code is not manipulated or changed inadvertently. Availability - ensuring that data is available to the right person at the right time.

Data processed in the TSD requires a high degree of confidentiality and integrity. Availability is also important in the sense that the right person shall have access to the right data, but minor focus will be put on ensuring that solutions has an uptime of 24x7. This may change in the future, if it will be decided that the TSD shall host clinical data (e.g. patient data etc). But such a change would require a reconsideration and perhaps significant changes in some components of the infrastructure.

6. Security assessment

6.1 Method

The risk assessment is performed with the methodology used at the University Center for Information Technology (USIT). It follows the guidelines provided by UNINETT and is based on a collection of risk factors that are assessed based on the probability and consequence.

The risk assessment is carried out by evaluating risk elements that can threaten the confidentiality, availability or integrity of TSD. Risk elements are extracted through assessments made in the design and development phase of the solution and finally evaluated in the Council of Changes, a periodic round table with decisional power involving the technical advisors, the administrative leaders and the IT-security officer.

The present risk assessment methodology is an updated version of the previous one and it is aligned with UNINETT guidelines for risk assessments. Compared to the previous risk analysis conducted in 2017, the only significant infrastructural change has been the addition of the HTTP API. This has been factored into the risk analysis.

6.2 Members

The assessment is done by:

Espen Grøndahl – IT-Security officer
Gard Thomassen – Service owner
Leon du Toit – Service manager
Morten Werner Forsbing – Technical Coordinator

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017385

7. Acceptance Criteria

Security will always be a consideration and a balance between usability and security. The level of usability should be high enough that the possibility to choose a less secure alternative is not wishful. Some risks will therefore be acceptable to get the functionality that users need.

However given the high security profile of the solution very few risks are considered acceptable. There must be low risk associated with inadvertently exposure of sensitive data. If there are security breaches, confidentiality and integrity are prioritized over availability. This will be emphasized in the risk assessment.

7.1 Evaluation Scales

Risk elements are evaluated on a scale from 1-4, probability, and 1-4 in consequence, where 1 is associated with low probability, or little or no consequence and 4 is very likely and very severe consequence.

In detail, this security assessment adopts the scale suggested by UNINETT to evaluate the probability:

Low (1)	Medium (2)	High (3)	Very High (4)
One time every 10 years or more seldom	One time per year or more seldom	One time per month or more seldom	Weekly

In details, this security assessment adopts the following scale to evaluate the consequences:

	Personal Data	Project Owner	Service Provider (USIT)
Very Severe Consequences (4)	Incident involving unjustifiable lack of security for personal data	- Incident involving loss of data or communication of data to unauthorized parties - Incident leading to irreparable financial losses	
Severe Consequence (3)		- Incident involving loss of data - Incident leading to significant financial losses	Incident leading to substantial and irreparable financial loss or serious loss of reputation / integrity.
Moderate Consequences (2)		Incident involving service with inadequate quality and low availability	Incident leading to substantial financial loss that can be recovered or loss of reputation / integrity due to

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017385

			compromising of infringing information
Little or no Consequences (4)			incident involving loss of trust between the project owner and service provider

Each risk element is associated with a value obtained by multiplying the values associated with the probability and the consequences.

7.2 Risk elements between 1 and 3:

Risk elements between 1 and 4 are considered acceptable either as they are or after the implementation of dedicated measures and/or routines.

7.3 Risk elements between 4 and 8:

Risk elements between 5 and 8 must be evaluated carefully. They might be considered as acceptable for a short period of time, while the necessary mitigating safeguards are being planned.

7.4 Risk elements between 9 and 16:

Risk elements between 9 and 16 are not acceptable. These risks require service interruption and/or must be compensated with manual control and strong routines until the risk is reduced.

8. Description and assessment of the risk elements

Risk elements are listed in the attached excel file. They are numbered with serial number. Here is a description and assessment of elements requiring mitigating measures or special attention (category [4-8]).

Risk element 5

The risk for data leakage between projects as a consequence of intentional action has inevitably high consequences but low probability. The projects have a dedicated VLAN and no network traffic between project VLANs is allowed. In addition, the access to the resources on a given project is given through group policies. The implemented safeguards have been tested in a penetration test undergone by TSD in summer 2016 and conducted by a well-known security expert. The test was successfully passed as the attacker did not manage to leak data between projects.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017385

Risk element 7
The risk has high consequence but low probability. The logon to the TSD is behind two factor authentication and therefore requires valid username password and one-time code. In the rare case in which the username and password are guessed, the one-time code is impossible to guess.

Risk element 14
The risk that data are exported through mechanism other than the filelock is continuously evaluated by periodic penetrating testing and adequate safeguards are implemented as different scenarios are discovered.

Risk element 15
The risk of someone hacking the TSD API, thereby gaining access to data stored in memory, as it is being transferred over the network. This is mitigated by two factors: 1) unless in the project area, or in an administrative VLAN protected by the firewall, no data is kept in memory unless encrypted, and 2) the API architecture is such that the application server which has access to plain-text data, stored in the project, is protected by several layers and authentication mechanisms with high level of assurance.

Risk element 16
The risk of damages caused by disloyal sys-admins and technical staff members is reduced by increasing the awareness of the technicians with regards to the importance of the service and the consequences of its failure. In addition, every sys-admin uses private dedicated user credentials in TSD and most of the operations are traceable.

Risk element 19
The event of fire or similar unfortunate events might have severe consequences. Even if the backups of the disks are kept in a different physical location, the disruption of the machine room could result into severe damage of the service infrastructure and long outage. Existing safeguards consist of fire alarms and monitoring of the machine room. In addition, TSD is investigating a mechanism to make off-site replicas of the data repository and infrastructure.

Risk element 20
The service is located in a room behind three sets of doors, the first (more extern) requires a valid card to enter, while the second and third requires card and code. Both the card and the code are private to the technicians and are issued by the University of Oslo. The system is not locally controlled, but centrally controlled at UiO. To reduce the risk even further inadvertently TSD are working to implement an extra lock to secure the second innermost door. The entire area including machine-rooms are under video surveillance once entering the first door. However, the combined use of the code and the card significantly reduces the probability of a successful unauthorized access. The risk element, compared to the previous risk assessment has been downgraded to acceptable risk (was 6).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017385

9. Final assessment and follow up

Every point on the previous version of the risk assessment has been rectified in such a way that the risk has been reduced to an acceptable level.

10. Conclusion

According to our assessment, the most serious risk elements has been downgraded to acceptable risks. The solution has been designed from the start to have very high security standards and the present assessment has revealed no weaknesses or backdoors into the system.

TSD has also established a system for incident handling. Cases are reported and archived so that we can easily look at trends and prevent future similar problems. Serious discrepancies will be reported to the end users.

A Council of Changes (in Norwegian "Endringsråd") has been established to periodically evaluate the security aspects of the present set up and of on-going developments. The Council involves the IT-security officer, the service owner, the service manager, the technical coordinator and the security experts. Main customer representatives are invited, but veto lies with the IT-security officer. The Council supervises any system-changes and guarantee that the security standards are respected even when the changes are so small that they do not invoke a Risk Assessment upgrade.

11. Attachments

A. "Whitepaper TSD "
B. Risk elements TSD

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017385