PRECISE4Q

PREDICTIVE MODELLING IN STROKE

DELIVERABLE

Project Acronym: **Precise4Q**

Grant Agreement number: **777107**

Project Title: **Personalised Medicine by Predictive Modelling in Stroke for better Quality of Life**

D2.4 – Functional harmonization of Users, Roles and Access

Revision: 1.0

Authors and Contributors	Catalina Martínez Costa (CMC) from Medical University of Graz(MUG); Jose Antonio Miñarro Giménez (JAMG) from University of Murcia (UMU); Nikola Lazovski (NL) from QMENTA; Paulo Rodrigues (PR) from QMENTA			
Responsible Author	Catalina Martínez Costa		**Email**	catalina.martinez@medunigraz.at
	Beneficiary	MUG	**Phone**	+4331638517880

Project co-funded by the European Commission within **H2020-SC1-2016-2017/SC1-PM-17-2017**		
Dissemination Level		
PU	Public, fully open	X
CO	Confidential, restricted under conditions set out in Model Grant Agreement	
CI	Classified, information as referred to in Commission Decision 2001/844/EC	

Revision History, Status, Abstract, Keywords, Statement of Originality

Revision History

Revision	Date	Author	Organisation	Description
0.1	17/10/19	JAMG	UMU	Draft of the functional analysis of Partners and repositories according to PRECISE4Q work plan to define User and Roles for accessing the Data Warehouse
0.2	28/10/19	CMC	MUG	Review of the draft
0.3	29/10/19	NL	QMENTA	Security and privacy description of the Data Warehouse in QMENTA platform
0.4	29/10/19	JAMG	UMU	Final version
1.0	29/10/19	CMC	MUG	Final review

Date of delivery	Contractual:	30/04/19	Actual:	31/10/2019
Status	final ☒ /draft ☐			

Abstract (for dissemination)	This document presents the functional analysis of PRECISE4Q project in order to define the Users/Roles. We have conducted an analysis of the tasks' assignments of each project partner and which permissions are required according to the repositories defined in the Data Warehouse. Thus, we include two tables where the first table describes the interrelationships between partners, repositories and tasks of the work plan, and the second table describes the interrelationships between the partners, the repositories and the required permissions according to the description of the assigned tasks.
Keywords	Access Control; Users/Role Permissions; Functional Analysis

Statement of originality

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Table of Content

List of Tables

Executive Summary

This document describes the interrelationships between the tasks defined in the project, the repositories of the Digital Stroke Patient Platform data warehouse, the project partners, and the different user/role permissions. We provide two tables, the first one describes the correlations between the particular tasks where each partner access the content of the repositories; the second table describes for each repository and permission (READ, WRITE and CREATE) the partners that require to be granted.

These tables justify the required permissions for each partner and repository based on the tasks assigned to them in the Work Plan. These permissions may change due to change in the description of the tasks. In this sense, the amendment of Work Package 2 (Data Management) implies QMENTA will be responsible for managing user/role permission of the Data Warehouse. This role is described in the deliverable, but it is out of the analysis of access level for each partner. Therefore, QMENTA will have two roles in the Data Warehouse, the manager role (full access to the Data Warehouse) and task-related role. The section 5 of the document describes the security measurements of the Data Warehouse to avoid any unauthorized access to the content of the repositories and, finally, the last section indicates the conclusions of the work.

1. Introduction

The project data warehouse is described in deliverables from 2.1 to 2.4. This document describes the functional requirements of each partner to interact with the project data warehouse based on the description of the Work Plan. Security and privacy measurements are implemented inside QMENTA platform which supports the Digital Stroke Patient Platform.

The functional permissions analysed are reading, writing and creating repositories for the project data warehouse. Reading will allow partners to access, search, filter and download the content of the repositories they are granted. Writing will allow partners to upload and modify the content of the repositories they are granted, i.e. adding or modifying lines in the repository. Finally, creating permission will allow partners to modify the schema of the repositories, i.e. adding new columns or restricting the values of the columns.

The following sections will describe in detail the repositories defined in the Digital Stroke Patient Platform data warehouse; the roles of each partner in the project; the functional analysis of the interrelationships between partners, tasks, repositories and permissions required; a brief description of how users and roles are managed in QMENTA platform; and the conclusions. The access to the data will of course be restricted according to the approval of the corresponding ethics boards (cf. deliverable 2.1).

2. Data Warehouse Repositories

In the PRECISE4Q Data Warehouse, we have identified several repositories according to the input datasets and output results associated with the tasks in the work plan. Most of these repositories are described in the work package 2 - Data management, work package 3 - Data semantics and harmonization and work package 4 - Predictive modelling. In particular, we defined the repositories Raw datasets (RD), NLP resources (NLP), Integrated and Harmonized data (IH), Machine Learning modelling (MLm) and Machine Learning prediction (MLp).

The RD repository contains the clinical datasets provided by the project partners GUT, LIU and UOT. These datasets use different data formats and can be structured and unstructured. Due to strict local security requirements the clinical datasets provided by AOK will not be part of Digital Stroke Patient Platform. However, CUB will be allowed to access it through a local system provided by AOK. The description of the content of these datasets is not the goal of this document but of deliverable D2.1.

The NLP repository contains the annotations and information extracted from RD repository. The schema of the NLP repository uses a common ontology that partially integrates the extracted information from different datasets. However, the data can include different levels of granularity, value units, etc.

The IH repository contains the standardized representation of integrated and harmonized data from NLP repository, including relevant data from external dictionaries and ontologies. The schema of the IH repository follow the same ontology as NLP repository but data is transformed based on the agreement between partners to represent the relevant information for training and learning predictive models about the four stroke healthcare stages defined in the project: (1) prevention setting; (2) acute stroke treatment setting; (3) rehabilitation setting; and (4) reintegration setting.

The MLm repository will store training datasets as well as the different versions of the machine learning models generated for each stroke healthcare stage described above. This repository will be used to develop the prediction tools and clinical decision support systems.

Finally, the MLp repository will store the validation data and the results of the prediction with the different versions of the ML models in MLm repository.

3. Partners effort

There are 12 partners in PRECISE4Q consortium: Charité University Medicine Berlin (CUB), Empirica Communication and Technology Research (EMP), Dublin Institute of Technology (DIT), University of Zurich (UZH), University of Tartu (UOT) Guttmann Institute Neurorehabilitation Hospital (GUT), Linkoping University (LIU), Medical University of Graz (MUG), University of Murcia (UMU) (upon amendment), German Research Center for Artificial Intelligence (DFKI), Public Health Insurance for Northeast Germany (AOK), QMENTA (QMENTA). The contribution that each partner provides to the project varies and is described in the work plan. Thus, the interrelationships between partners and repositories will depend on their tasks assigned within the work plan. Table 1 contains the correlations between the partners, the repositories described above and the tasks from PRECISE4Q project work plan. In order to create Table 1, we have defined the following criteria:

- We do not include tasks from work package 1 - "Patients' Needs and Ethical Framework" because it does not need direct access to the clinical datasets but most of the partners will access it from other tasks.
- We do not include tasks from work package 7 - "Project Management, Communication and Outreach" because its main goal is to support the project according to the work plan, including administrative, financial and legal issues. Therefore, direct access to the clinical data is not required.
- The assignment of tasks between partners and repositories may differ from the work plan in order to restrict the access to the partners that will effectively use the clinical data. So, we use the following guidelines:
    - We selected only the partners associated with tasks where they have a relevant contribution that requires actual access to the data.
    - The contribution depends on the effort assigned to them in the work package and their expertise. Partners with low rate of person/months (PMs) assigned in the work package are usually not accessing data.
    - We have also included partners that have manifested their interest to access the data regarding their contribution to achieve the goals of the tasks and work package.

	Partner	**RD**	**NLP**	**IH**	**MLm**	**MLp**
1	UMU	T2.2, T3.1, T3.2, T3.6	T2.2, T3.1, T3.2, T3.3, T3.6	T2.3, T2.6, T3.2, T3.4, T3.5, T3.6, T3.7, T4.1, T4.3		
2	MUG	T2.3, T2.2, T3.1, T3.2, T3.6	T2.2, T2.3, T3.1, T3.2, T3.3, T3.6	T2.3, T3.2, T3.4, T3.5, T3.6, T3.7	T2.3	T2.3
3	LIU	T2.1, T2.2, T3.2, T3.6	T2.2, T3.2, T3.6	T3.2, T3.4, T3.5, T2.7, T3.6, T3.7, T4.1, T4.2	T2.7, T2.8, T4.2, T4.3, T4.4, T4.5, T4.6, T4.7, T4.8, T4.9, T4.10	T2.7, T2.8, T4.2, T4.3, T4.4, T4.5, T4.6, T4.7, T4.8, T4.9, T4.10, T6.1

4	DFKI	T2.2, T3.1, T3.3, T3.4, T3.5, T3.6	T2.2, T3.1, T3.3, T3.4, T3.5, T3.6	T3.4, T3.5, T3.6, T3.7		
5	DIT			T2.7, T3.6, T4.1, T4.2	T2.7, T2.8, T4.2, T4.3, T4.4, T4.5, T4.6, T4.7, T4.8, T4.9, T4.10	T2.7, T2.8, T4.2, T4.3, T4.4, T4.5, T4.6, T4.7, T4.8, T4.9, T4.10, T6.1
6	UOT	T2.1, T2.2, T3,3	T3.3	T2.7, T3.6, T4.2, T5.2	T2.7, T2.8, T4.2, T4.3, T4.5, T4.10, T5.2	T2.7, T2.8, T4.2, T4.3, T4.5, T4.10, T5.2
7	GUT	T2.1, T2.2		T3.6, T5.1, T5.4, T5.5	T4.7, T4.8, T4.9, T4.10	T4.7, T4.8, T4.9, T4.10, T5.4, T5.5, T6.1, T6.2
8	CUB			T2.7, T3.6, T4.2, T5.1, T5.2, T5.3, T5.4	T2.7, T2.8, T4.2, T4.3, T4.4, T4.5, T4.6, T4.7, T4.8, T4.9, T4.10, T5.2	T2.7, T2.8, T4.2, T4.3, T4.4, T4.5, T4.6, T4.7, T4.8, T4.9, T4.10, T5.2, T5.4, T6.1, T6.2, T6.3, T6.4, T6.5
9	QMENTA	T2.3, T2.6	T2.3, T2.6	T2.3, T2.6, T2.7	T2.3, T2.6, T2.7, T2.8, T4.6, T4.10	T2.3, T2.6, T2.7, T2.8, T4.6, T4.10, T6.1, T6.2, T6.4, T6.5
10	AOK	T2.1, T2.2		T4.1, T5.1	T4.5, T4.10	T4.5, T4.10, T6.1, T6.2, T6.3, T6.4, T6.5
11	EMP					T6.1, T6.2, T6.3, T6.4, T6.5
12	UZH			T3.7, T4.1		

Table 1 Interrelationships between partners, repositories and tasks from the work plan of PRECISE4Q project

4. Functional analysis of user/roles

We use three different levels of permission to access the repositories in the Digital Stroke Patient Platform Data Warehouse: READ, WRITE and CREATE/ALTER. Partners that are granted with READ permission will be able to search, filter and download the content of the repository. Usually, the partners with this permission will retrieve the content produced by other partners, process it and store their results in a different repository of the Data Warehouse. Partners that are granted with WRITE permission, have also the READ permission, and are able to store and modify the content of the repository. This permission will allow partners to store their results obtained in their assigned tasks. CREATE/ALTER permission allows to create the schema of a repository or modify their content. This permission is different from WRITE since it does not allow to change the content of a repository but to define the schema that every partner accessing the repository should follow.

As described in Table 1, QMENTA is in charge of hosting, setting up and managing the Data Warehouse (Task 2.6). This means that QMENTA has all administrative permissions to create users, grant permissions and access and modify all repositories in the Data Warehouse.

Table 2 shows the correlations between the repositories described in section 2, the three levels of permissions described above, and the partners of the project described in section 3. The administrative role of QMENTA for managing the Data Warehouse is not represented in Table 2.

Repository	READ	WRITE	CREATE
RD	UOT, GUT, LIU, DFKI, MUG, UMU, AOK	UOT, GUT, LIU	UOT, GUT, LIU
NLP	DFKI, MUG, UMU, LIU, UOT	DFKI, MUG, UMU, UOT	UMU, DFKI, UOT
IH	UMU, LIU, MUG, DFKI, UOT, GUT, QMENTA, DIT, CUB, UOT	UMU, LIU, DFKI	UMU, LIU, DFKI
MLm	DIT, LIU, CUB, UOT, QMENTA, AOK, GUT	DIT, LIU, QMENTA	DIT, LIU, QMENTA
MLp	DIT, LIU, CUB, UOT, QMENTA, EMP, AOK, GUT	DIT, LIU, CUB, QMENTA	DIT, LIU, CUB, QMENTA

Table 2 Functional analysis of the permissions required by the PRECISE4Q partners to access the repositories in the Digital Stroke Patient Platform Data Warehouse

5. Security and privacy description of the Data Warehouse in QMENTA platform.

As it is described above the access control of the Data Warehouse is managed by QMETNA. The user and roles are defined at repository level. Therefore, it is defined for each partner and repository a different user with the permissions they required according to Table 2.

Within the QMENTA platform, all data are organized in projects. Each project is of certain type, which means that they might differ in features, roles, and in available modules provided to the user. Each

user must be assigned one or more roles when added to a project. For each project type the system provides a list of predefined roles allowed to be assigned.

A role is a set of permissions assigned to that role. Each permission is described by an action (e.g. READ), an object/entity over which the action is performed (e.g. FILE) and a set of additional object constraints. Object constraints allow fine-grained access control in which a specific group of objects can be targeted. For example, if a role has the permission (FILE, READ, OWNER), this means that a user with this role can only read the files owned by the user and no other files created by other users. FILE refers to an entity, READ refers to an action, and OWNER refers to an additional constraint.

Upon each request, the system uses the request token to identify the user and then it loads and calculates a summary permission list for the appointed project in the request (a "summary" because the user can be assigned several roles that overlap in certain entities and actions). Together with the token, the user sends a URL of the service endpoint. On another side, each service endpoint specifies a list of permissions the user has to have in order to be allowed that service. Therefore, if the requested permissions (by the endpoint) are not found within the calculated permission list, the user will be denied the usage of that service. This way the access list can vary from one project context to another: when defining the permissions, the administrator can define the context to which they apply. Currently, within the QMENTA system two contexts levels are supported: general and project level context.

Regarding the P4Q, each repository might be defined as a separate project. Different roles can be assigned in different projects, and all the involved users will be divided into groups depending on the institutions they are coming from. These user groups can be included into the permissions definition, thus reflecting the access matrix described in the previous section of this document (See Table 2)

6. Conclusions

In this deliverable we have described the functional analysis of the work plan in order to discover the needs of the project partners regarding the users and roles permissions to access the Data Warehouse. We have defined three levels of permissions (READ, WRITE and CREATE) and we have divided the content of the Data Warehouse into five independent repositories (RD, NLP, IH, MLm, MLp). Thus, the permissions are assigned to each partner and repository according to the description of the tasks in the work plan.

The functional analysis of the tasks shows three major phases: (1) data collection phase that includes a description of the content of the datasets and the input format of data sources; (2) data preparation phase that includes extraction, integration, harmonization and semantics tasks; and (3) modelling and prediction phase that applies machine learning methods for modelling and prediction tasks). This division is also reflected in the division the repositories, where RD is related to data collection phase, NLP and IH are linked to data preparation and MLm and MLp are related to modelling and prediction phase.

As a result, in Table 2 we show for each repository the level of access we will provide for each partner. As a consequence, the content of this table will be reflected in the definition of user/roles of the Data Warehouse.