DELIVERABLE

Project Acronym: **Precise4Q**

Grant Agreement number: **777107**

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

D2.2 – Implementation of accessing and bridging functionalities

Revision: 1.0

Authors and Contributors	Paulo Rodrigues (QMENTA); Nikola Lazovski (QMENTA); Catalina Martínez Costa (MUG); Jose Antonio Miñarro Giménez (UM)			
Responsible Author	Paulo Rodrigues		**Email**	paulo@qmenta.com
	Beneficiary	QMENTA	**Phone**	+34 933 282 007

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

Revision History

Revision	Date	Author	Organisation	Description
0.1	16/10/19	Paulo R	QMENTA	Initial Draft
0.2	20/10/19	Jose G	UM	Introduction
0.3	25/10/19	Nikola L	QMENTA	QMENTA functionality description
0.4	29/10/19	Paulo R	QMENTA	Review
0.5	30/10/19	Catalina MC	MUG	Review
1.0	30/10/19	Paulo R	QMENTA	Final Review

Date of delivery	Contractual:	30.04.2019	Actual:	31.10.2019
Status	final **x** /draft ☐			

Abstract (for dissemination)	This document outlines how the Cloud Platform is used to provide a Data Warehouse to store, retrieve and model data from various the data sources.
Keywords	platform, stroke, integration

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This deliverable contains original unpublished work except where clearly indicated otherwise.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.

List of Figures

Executive Summary

For data exchange between organizations, the QMENTA Cloud Platform from QMENTA will be provided as data infrastructure to access, search, filter, retrieve, and store available data (structured and unstructured). Regarding integrity and the technical security, the QMENTA infrastructure already provides a secure IT system that is used for clinical services. QMENTA will provide the procedures for importing and exporting data from the QMENTA Cloud Platform. The stroke data will be organized in different data repositories, and detailed role/permission models will be applied in order to enable fine-grained data access.

All partners developing tools to analyse the stroke data can use the QMENTA API and QMENTA SDK to run analysis locally or within the QMENTA platform, thus generating valuable knowledge that is saved back into the QMENTA data repository.

1. Introduction

This deliverable aims at describing the functionalities to import the datasets, and to access and modify their content in the Digital Stroke Patient Platform. The content of the data warehouse is divided in several perspectives: (1) Raw datasets, as they are provided by each data provider partner; (2) text analyzed datasets that contains the results of applying NLP tools to extract relevant information from clinical free text; (3) integrated and harmonized datasets, that provide an expanded representation of the datasets with formal representation of their semantics and a common level of harmonization to facilitate sharing them; (4) Machine Learning modeling/predicting datasets. Each perspective will be associated with different repositories in the data warehouse where permission and roles to access and modify their content will be managed at a low granularity level. Moreover, having independent repositories for each perspective will facilitate decoupling the tasks from different partners in the different Work Packages (WP).

In this deliverable, the description of the Digital Stroke Patient Platform related to the access and exploitation of the datasets is provided. It is closely related to the content of deliverables D2.1, D2.3 and D2.4. The content and format of the raw datasets are described in D2.1. The overall data warehouse for sharing and storing the datasets is described in D2.3. The description of Users, Roles and Access is provided in D2.4.

In the next section (section 2) we describe the QMENTA platform which supports the Digital Stroke Patient Platform and provides the infrastructure to access, search, filter, retrieve, and store available datasets. In section 3, we describe how the data warehouse of the project is integrated into QMENTA platform. In section 4, we describe how the project partners will interact with the data warehouse.

2. Functionalities for exploiting datasets of Digital Stroke Patient Platform.

The Digital Stroke Patient Platform is integrated into the QMENTA platform and, therefore, it uses the functionalities provided by this platform for accessing, searching, filtering, retrieving and storing the available datasets. The project data warehouse is part of the Digital Stroke Patient Platform and allow partners to share and process the clinical datasets.

The partners that provide the raw datasets (UOT, LIU, GUT) will export their repositories into the platform following a custom data model related to the schema of their repositories. The custom data model will follow the description of the datasets from D2.1. Thus, the uploaded datasets are available in the platform for other partners. The data warehouse implements strict security measures and user profiles with different permissions for each partner to prevent any unauthorized access or modification of the content of the repositories.

The raw datasets will be processed by the text analysis and data integration and harmonization modules within the project data warehouse. QMENTA platform will provide accessing and storing data warehouse functionalities as well as an API to allow the specific modules integrated into the data warehouse to access, retrieve and modify the data.

Create an account & Login

In order to sign up, the user has to visit the platform website: https://client.qmenta.com/#/register where the user is required to enter his/her name, email and password in the registration form. After completing the form, a meeting has to be scheduled with QMENTA before the account gets activated by the platform administrator.

After activating the account, the user can visit the page https://client.qmenta.com and login into the system. The system blocks any account after 5 unsuccessful consecutive login attempts.

Main working area

After a successful login, the user is presented with the main working area of the system. The main parts are explained in the following figure:
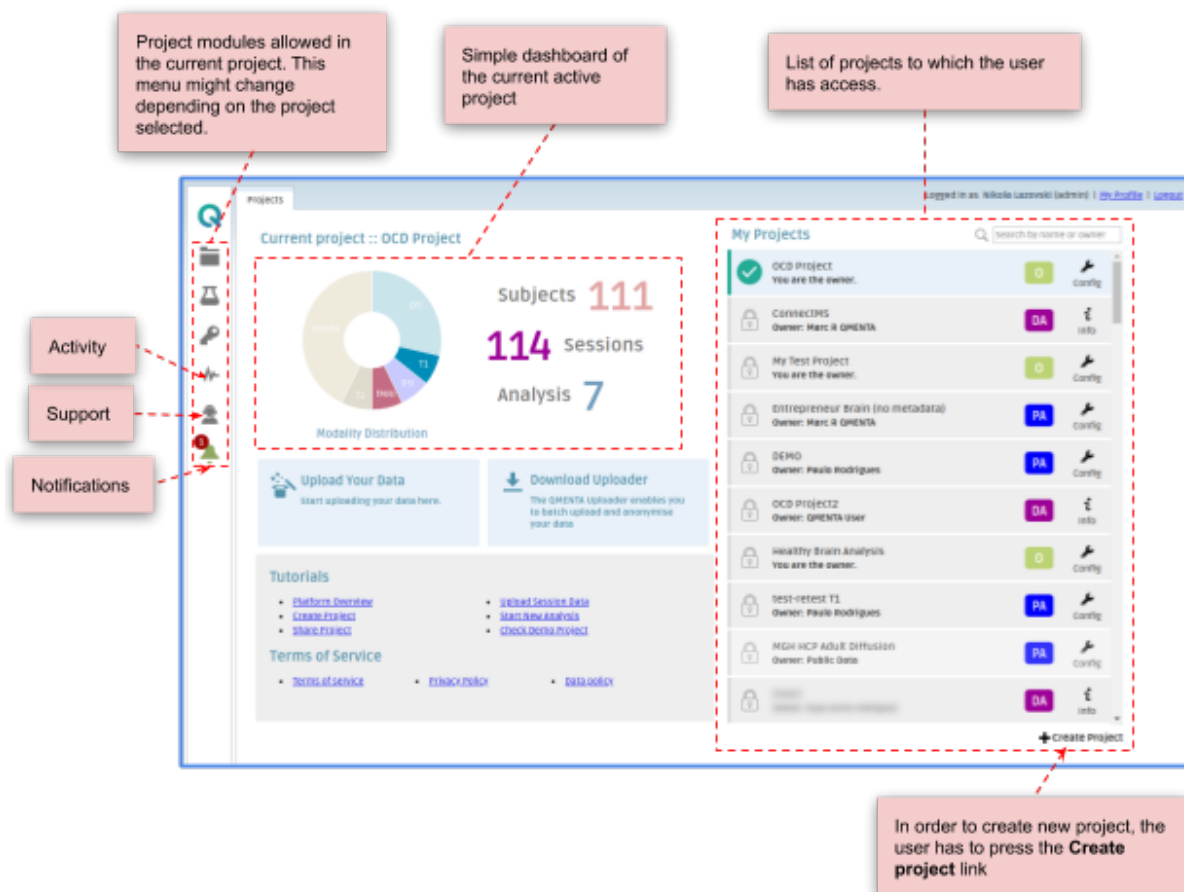
Figure 1 Main working area of the QMENTA Platform

Each project is of a particular type and that means that depending on the type, users are provided with different roles and modules. The most common modules among the project types are:

- **My Data**: In this module the user can upload and alter session data, define metadata fields, and perform advanced search over the session data.
- **My Analysis**: This module allows users to run various analysis and to visualize results.

Data organization and advanced search

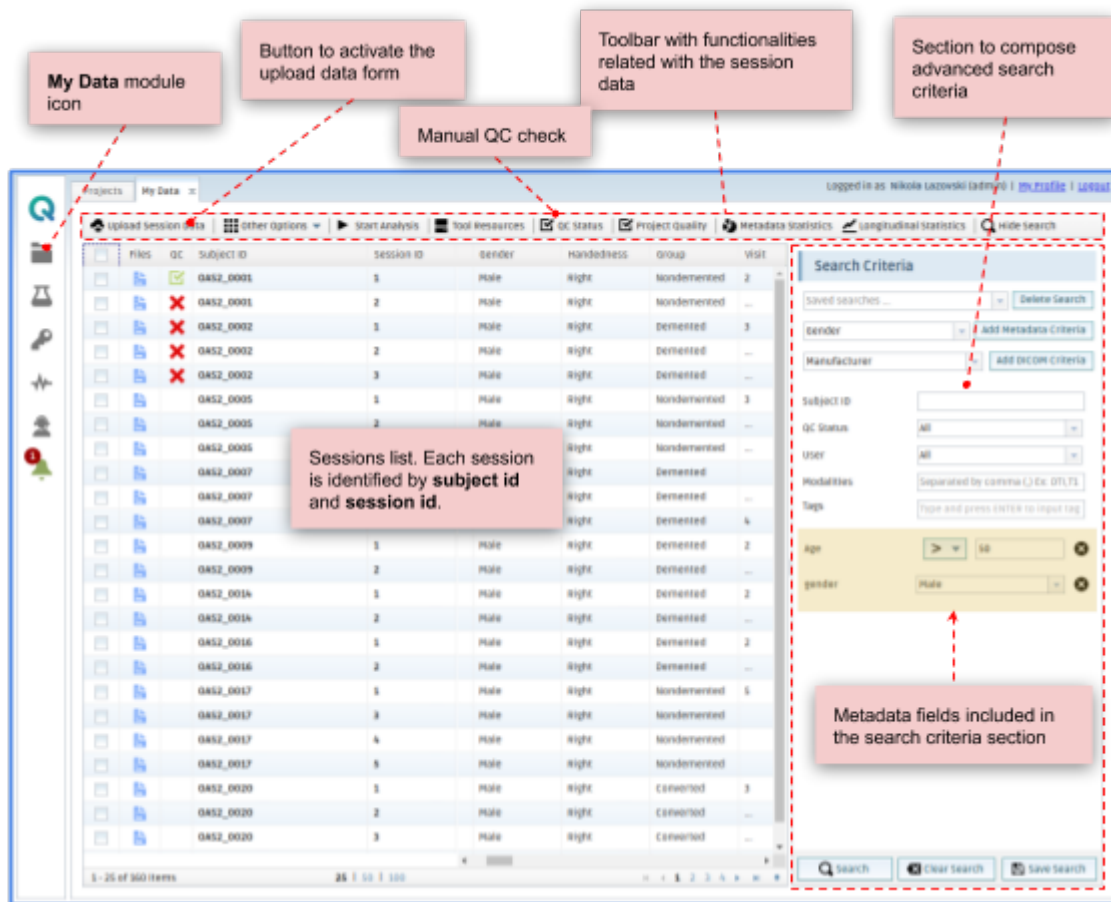The My Data module is one of the most common modules for data organization (See Figure 2):

Figure 2 My Data module

Uploading data

To upload a session, the user has to click the Upload Session Data button. In the presented form, the user chooses the ZIP file containing all session files, and then populates the mandatory fields marked with a red star. After clicking the Save button, the ZIP file is being uploaded and the QMENTA platform automatically starts an analysis that analyses, labels and stores the files of interest. Note that this type of analysis can be chosen on a project level so that different tools can be used in different projects to perform this work.

Figure 3 Form to upload session data

In order to define the metadata fields and the session types, the user has to use the Metadata Manager located in the Other Options submenu.

Downloading Data

The session data can be downloaded mainly in three ways:

- **Downloading particular files**: The user has to right-click the session row and choose Show files option. In the presented form, the user has to select the desired files and to press the Download button. If there are more than one files selected to be downloaded, the system downloads them all compressed in one ZIP file.
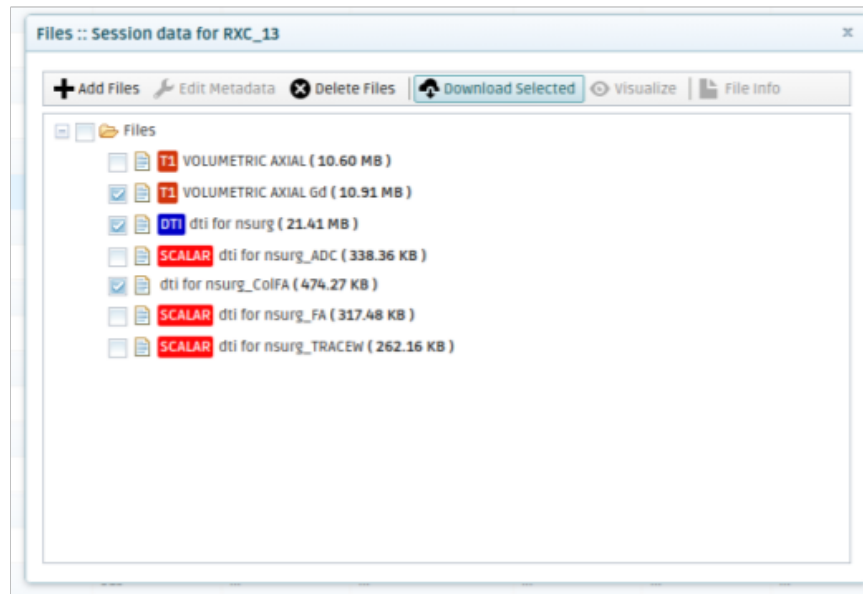
Figure 4 Listing of session files

- **Download File Manager:** This module is used to download more files that satisfy complex criteria, including the current Advanced Search Criteria. All files of interest are downloaded as a ZIP file in which the data are organized by sessions and by analysis.
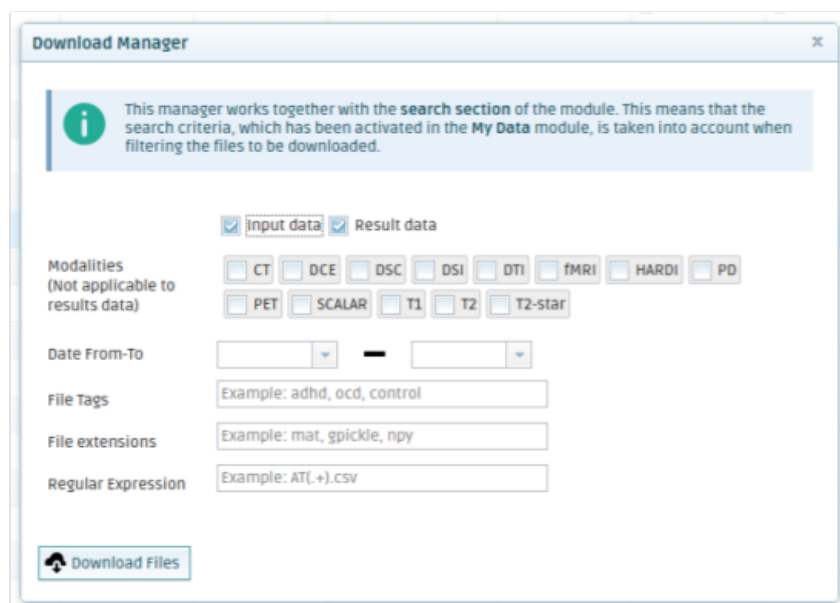

Figure 5 Download File Manager form

- **Export CSV**: Apart from files, a user can download all metadata fields and their values by using the Export CSV functionality located in the Other Options submenu.

Analyzing data

The My Analysis module serves to manage the analysis in a project. This module, depending on the user's permissions, offers various functionalities, from starting an analysis to observing analysis and visualizing analysis results. The system distinguish between several kinds of analysis:

- Standard fully automatic analysis,

- Semi-automatic analysis where user interaction is required, and

- Workflows: an analysis that is composed of other existing analysis.
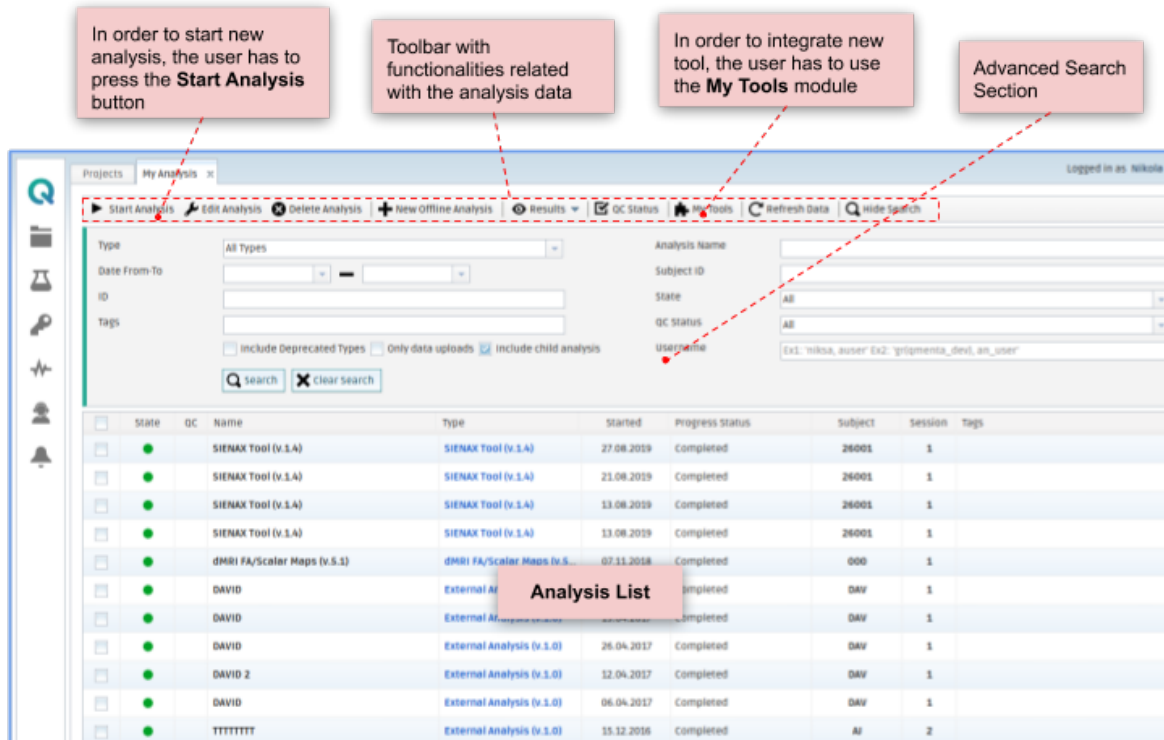

Figure 6 My analysis module

In order to start an analysis, the user has to click the Start Analysis button. In the presented form the user can choose the type of the analysis to run and can tweak the configuration options specific to each analysis (See Figure 7)

Figure 7 Form to start new analysis

When the analysis is successfully started the user can observe the state of the analysis, and when finished, the results can be downloaded and visualized.

Integrating tools

Partners can develop and run their own tools in two ways:

- They can develop their own scripts and by including the QMENTA Core library they can communicate the platform the same way the client web application does. Currently only a Python version of this library is provided. More information on how to install and use this Python library can be found on the following link https://docs.qmenta.com/core/

- They can analyze their data directly in the QMENTA platform by creating tools using the QMENTA SDK library.

QMENTA SDK

The QMENTA platform provides a unified solution for data management, processing and visualisation. The analysis (neuroimaging tools) are executed on the hybrid cloud infrastructure as Docker containers and treated as black boxes that communicate with the platform using our SDK.
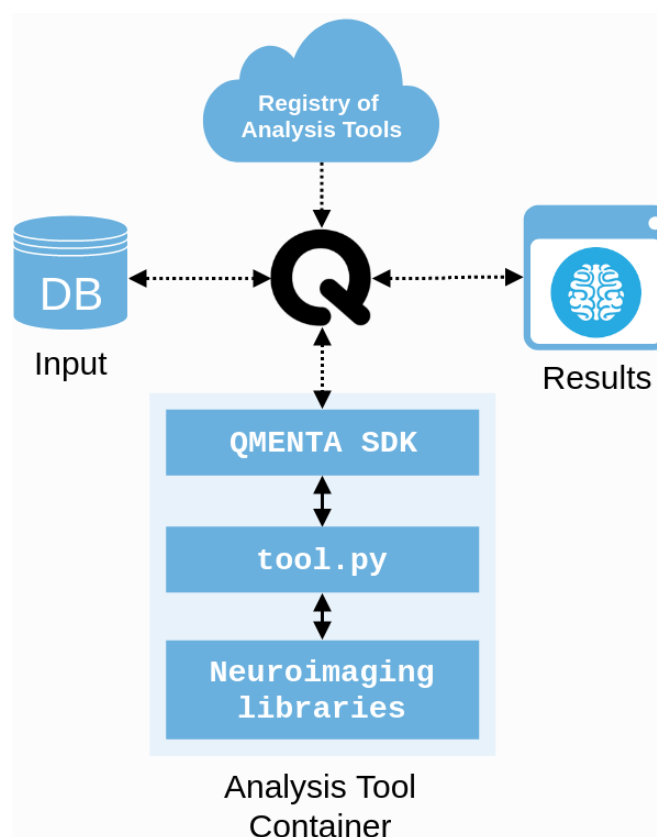

Figure 8 SDK usage with regards to QMENTA Platform

A simple Python script connects the QMENTA platform with any third-party program. This typically involves fetching the input data, starting the binaries or scripts and uploading the result files in a compartmentalized, traceable and secure computing environment.

Any data uploaded to the platform can be used for analysis with a simple step, and the resulting files can be managed and visualized within the same platform website.

In addition, tools can also be seamlessly customized to include external parameters using GUI elements that provide an easy way for users to choose between different options and tune the execution flow of the analyses. More information on how to integrate external tool is provided on the following link https://docs.qmenta.com/sdk.

3. QMENTA platform and Digital Stroke Patient Platform integration.

QMENTA platform provides a cloud based solution for storing and data analysis. The Digital Stroke Patient Platform is integrated in QMENTA platform and uses their functionalities to perform the corresponding text analysis, data integration, semantics and harmonization tasks, machine learning modelling and prediction. These tasks can be executed inside the scope of QMENTA platform using their API. Figure 9 shows how the Digital Stroke Patient Platform is integrated in QMENTA platform. Thus, the project data warehouse is deployed inside QMENTA platform and provides the functionality to search, retrieve, modify, upload and download, particular content of the repository. The Security Framework controls the external and internal access to the content of the data warehouse. The Digital Stroke Patient Platform consists of several modules developed in this project related to Natural Language Processing (NLP), integration and harmonization, learning predictive modeling, and semantics tasks.

At a later stage, the Digital Stroke Patient Platform will be separated as a full, easily portable, service also holding the data: that way it can be easily integrated or run independently.
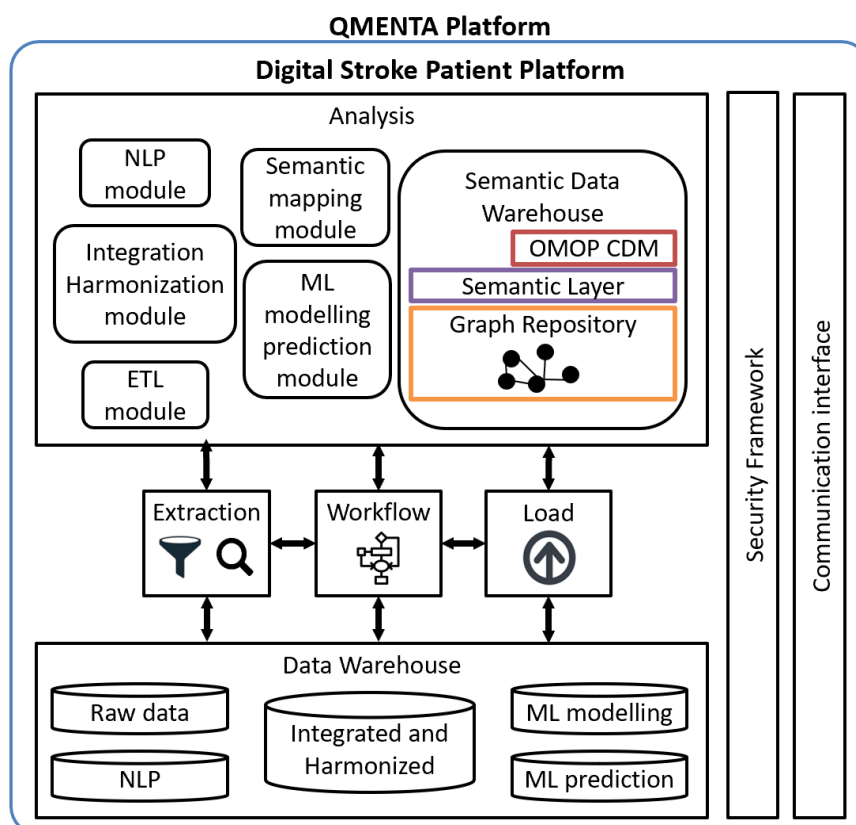

Figure 9 Integration of the Digital Stroke Patient Platform into QMENTA platform

All partners within this project, by using the SDK or the QMENTA API, as explained earlier can build their own tools. Those tools can use the QMENTA API to fetch data in order to derive more knowledge, which is also saved into the corresponding data repositories. For training purposes, the machine

learning/predictive Modelling will interact with the data repository which has to be enriched by semantic annotation of the raw data, imported from different sources. For clinical purposes, the "prediction knowledge repository" will be linked to a subset of the clinical data to map the predictive modelling to a concrete, specific clinical case.

4. Conclusions

The QMENTA Cloud Platform provides the data infrastructure to access, search, filter, retrieve, and store available data (structured and unstructured). The stroke data will be organized in different data repositories, and detailed role/permission models will be applied in order to enable fine-grained data access.

All partners developing tools to analyse the stroke data can use the QMENTA API and QMENTA SDK to run analysis locally or within the QMENTA platform, thus generating valuable knowledge that is saved back into the QMENTA data repository.