PNC
Piano nazionale per gli investimenti complementari al PNRR
Ministero dell'Università e della Ricerca

D7.1

DATA MODEL AND STANDARD SPECIFICATION

Piano Nazionale Complementare (PNC) – Decreto Direttoriale n. 931 del 6 giugno 2022 – Avviso per la concessione di finanziamenti destinati ad iniziative di ricerca per tecnologie e percorsi innovativi in ambito sanitario e assistenziale

Project identifier: PNC0000007

Start date: 01/12/2022

Duration: 44 months

Website: www.fit4medrob.it

Editors:
Riccardo Bellazzi (Università di Pavia), Manuela Chessa (Università di Genova), Giuseppe Prencipe (Università di Pisa)

Due date of deliverable: 31/05/2023
Actual submission date: 26/05/2023
Date of resubmission: 13/05/2024
Version: 1.1

DISSEMINATION LEVEL OF DELIVERABLE		
PU	Public, fully open, e.g. web	X
CO	Confidential, restricted under conditions set out in Partners Agreement	

HISTORY OF CHANGES

VERSION	SUBMISSION DATE	CHANGES
1.0	26 May 2023	
1.1	13 May 2024	Added Introduction

TABLE OF CONTENTS

1 EXECUTIVE SUMMARY

Fit4Medrob requires the definition of a suitable ICT infrastructure able to support the implementation of robotics interventions in at least three (possibly combined) operational settings:

1) Hospital and ambulatory rehabilitation services
2) Nursing homes, protected care services
3) Home

In the rehabilitation context, the infrastructure should be designed to support the three main rehabilitation activities:

a) Delivery of the robotic intervention
b) Monitoring and assessment of the robotic intervention
c) Prescription of the robotic intervention

The high-level view of the data technological infrastructure of Fit4MedRob is thus represented by the components that must allow implementing the above-mentioned three activities:

a) **Delivery of the robotic intervention**:
    a. the robotic platform and its on-board data acquisition, processing and transmission components;
    b. a virtual environment to support the actuation of patients' rehabilitation activities, including exergames.
    c. a sensing system able to detect patients' activity and psychophysical state.

b) Monitoring and assessment of the robotic intervention:
    a. a data repository to store all monitoring data collected by the sensing system and the virtual environment;
    b. a health care-provider clinical dashboard to support data integration, analysis, and decision support, including patient's feedback and coaching recommendations. The dashboard should be integrated with the prescription system.
    c. In a telemedicine setting, a mobile communication platform, which may be integrated with the virtual environment, is also required.

c) **Prescription of the robotic intervention:**
    a. a prescription system able to collect patients' data from the Electronic Health record, check available guidelines and local protocols, and define the PRI (Progetto Riabilitativo Individuale) and the pri (programma riabilitativo individuale) that includes the robotic intervention. The prescription tool should be integrated with the EHR / Hospital information system as well as with the dashboard for data analytics.

The system should be based on a standardized data communication system to collect sensing data and transmit them in an interoperable format.

Overall, the purpose of this document is twofold: first, describe a general Fit4MedRob ICT architecture, in all of its components, that fit best the needs of a robotic rehabilitation infrastructure; second, it aims at giving a picture of the "local scenarios" (as of today and limited to the partners and experiences in the Fit4MedRob partnership) of the main adopted ICT infrastructures for handling robotic intervention and rehabilitation actions in hospitals/nursing and care services/homes.

2 Introduction

Fit4MedRob (Fit for Medical Robotics) project aims to revolutionize the current rehabilitation and assistive models for people of all ages with reduced or absent motor, sensory, or cognitive functions, by means of novel biorobotic and allied

digital technologies and of continuum of care paradigms. ICT technologies provide essential components in all phases of the deployment of robotic platforms in the context of the rehabilitation process, from prevention up to home care in the chronic phase. The goal of the deliverable is therefore to identify all ICT elements that are necessary for the implementation of the project by providing a general architecture that can be considered as a shared conceptual framework within the project, a description of its components and some use cases already available at the participating partners rehabilitation facilities. Deliverable D7.1 will also be the basis for the actual implementation of the pragmatic trials foreseen in Mission 1 of Fit4MedRob. The structure of the deliverable is as follows: The first section describes the overall architecture of the Fit4MedRob ICT infrastructure and details its main components and the functional workflow. Afterwards, a detail about the ICT components is provided for each of the phases of the workflow, including the delivery of the robotic intervention, its monitoring and assessment, and the prescription phase. A focus is then provided on the security and standards aspects as well as on inter institutional data sharing. Finally 5 use cases of running ICT infrastructures are reported.

3 The architecture of the Fit4MedRob ICT infrastructure

3.1 The components

At a high level, the Fit4MedRob ICT infrastructure will include several main components: a centralized control system, robust networking, sensor networks, cloud computing, mobile devices, and virtual and augmented reality technologies. These will enable the support needed to implement robotics interventions in the three main environments addressed by the project: hospitals, nursing homes, and homes effectively. In particular, such an ICT infrastructure should be composed by the following components (also refer to Figure 1):

1. A centralized control system composed by a Digital Support System (DSS) along with its dashboard: The ICT infrastructure should include a control system to manage and control the various robotics interventions in these scenarios. This system should (i) allow the definition and the remote monitoring (either synchronously or asynchronously to reduce the burden for the operator) of the programmed rehabilitation exercises, (ii) allow the remote access and control of the robots/sensors, (iii) enable authorized personnel to monitor and operate the equipment and the tools from a central location.
2. A robust networking system is essential for exchanging data and instructions between the centralized control system and the sensors/robots, and it is in general crucial to support the implementation of robotics interventions in healthcare settings. Also, a high-speed, reliable network is necessary to ensure the robots can operate efficiently and without interruption. Such a network should also keep into account all related security issues, by considering measures such as strong encryption, authentication, firewalls, intrusion detection and prevention systems, regular updates and maintenance, and network segmentation:
    a. All data transmitted over the network should be **encrypted** to prevent unauthorized access. This means that any sensitive information transmitted between the centralized control system, the robot/sensors, and other devices should be encrypted using strong encryption methods (note that this measure might not be necessary in case data is transmitted within an intranet or a private network)
    b. Access to the network should be restricted to authorized personnel only. This can be achieved through user **authentication** methods such as passwords, biometric identification, and multi-factor authentication.
    c. **Firewalls** can be installed to monitor and control incoming and outgoing network traffic. This can help prevent unauthorized access and mitigate potential attacks.
    d. **Intrusion detection** and **prevention systems** can be used to identify and respond to potential threats in real-time. These systems can detect and block unauthorized access attempts, suspicious activity, and other threats.

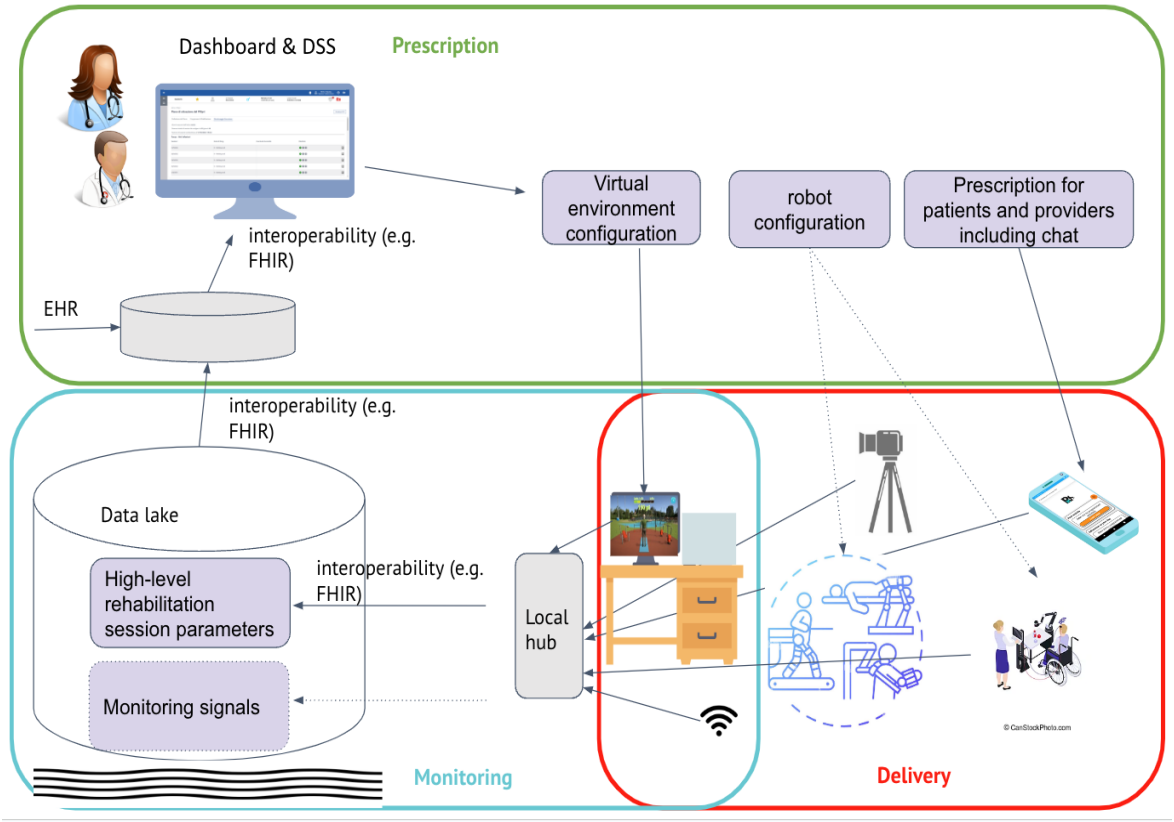


Figure 1. Overall ICT infrastructure's schema

e. **Network segmentation** can be used to separate different parts of the network into smaller, more manageable segments. This can help prevent unauthorized access to sensitive data and systems.

3. The delivery setting for rehabilitation: Sensors should be placed in strategic locations (such as, rehabilitation ambulatory, nursing homes, patient's home) to provide real-time data on the rehabilitation process and the overall activities of the robots. These sensors can include cameras, microphones, temperature sensors, and pressure sensors. Remote assistance of the subjects is allowed through a chatbot present on a piece of software (such as a mobile app) given to the subject, which allows not only to monitor the completed activities and the progress achieved, but also to receive reminders and advice for the next scheduled sessions.
4. Cloud Computing in EU infrastructures: this is another aspect of paramount importance within the robotic rehabilitation scenario: it can in fact be used to store and process the data collected by the robots/sensors. This can provide real-time analytics and insights, which can be used to optimize the performance of the robots and to improve patient outcomes. In more detail:
    a. Storage and Processing: Cloud computing provides a scalable and cost-effective solution for storing and processing the data collected by the sensors. This data can include information about the environment, patient activity, and robot performance. By using cloud storage, healthcare organizations can store large amounts of data without needing to invest in expensive on-site storage solutions. In particular, we exploit the concept of data lake, a centralized repository that allows organizations to store, process, and analyze large volumes of structured and unstructured data. In

healthcare systems; they can be used to store a wide range of data, including patient health records, clinical notes, medical images, sensor data, and other relevant information. Data lakes typically allow:

i. Data Integration: Healthcare systems generate a vast amount of data from various sources such as electronic health records (EHRs), videos, clinical notes, medical devices, and sensors. Data lakes allow for the integration of data from disparate sources, providing a comprehensive view of patient health that can be used to improve patient outcomes and enhance the quality of care.
ii. Scalability: Data lakes can scale to accommodate large volumes of data, making them well-suited for healthcare systems that generate significant amounts of data. This allows healthcare organizations to store and analyze data from multiple sources without worrying about storage capacity.
iii. Data Analysis: Data lakes enable advanced analytics and data processing techniques, such as machine learning and artificial intelligence, to analyze and extract insights from the data. This can help healthcare organizations identify patterns, trends, and anomalies in patient health data, enabling them to make more informed decisions and improve patient outcomes.
iv. Data Governance: Data lakes require robust data governance practices to ensure the data stored in them is accurate, up-to-date, and secure. This includes policies and procedures for data access, data quality, data privacy, and data security. Data governance practices also ensure compliance with regulatory requirements, such as HIPAA, which governs the storage and use of patient health data.
v. Collaboration: Data lakes can be shared across healthcare organizations, enabling collaboration between healthcare providers, researchers, and other stakeholders. By sharing data, healthcare organizations can improve the overall quality of care and accelerate research efforts.

b. Analytics and Insights: Cloud computing is useful for real-time analytics and insights that can help healthcare organizations optimize the performance of the robots and improve patient outcomes. For example, analytics can help identify patterns in patient activity, which can be used to support decisions about robot tasks and deployment.
c. Remote Access: Cloud computing provides remote access to the data and applications needed to operate the robots. Authorized personnel can access the centralized control system from anywhere with an internet connection, enabling them to monitor and operate the robots from a remote location.
d. Privacy: it is also essential to ensure that the cloud computing provider you choose is compliant with relevant regulations and standards, such as HIPAA and GDPR, and has appropriate security measures in place.

5. Virtual and Augmented Reality, and Exergames:
    a. Virtual reality (VR) technology can provide a simulated environment that allows users to interact with computer-generated objects and environments. In healthcare, VR can be used to provide training for personnel on how to operate and maintain the robots effectively. For example, healthcare workers can practice using the robots in a simulated environment before deploying them in the actual setting. VR can also be used to provide immersive experiences for patients, which can help reduce stress and anxiety.
    b. Augmented reality (AR) technology can be used to overlay computer-generated information onto the real-world environment. In healthcare, AR can be used to provide guidance to personnel as they operate the robots. For example, AR could be used to display instructions on how to perform a specific task or highlight potential hazards in the environment.
    c. Finally, exergames are video games that require physical activity to play. In healthcare, exergames can provide patients with a fun and engaging way to exercise and improve their physical health. For example, a robot could be used to play exergames with patients in a nursing home, encouraging them to stay active and engaged.

3.2 The functional workflow

In the following, we will better detail the **functional workflow** of the Fit4MedRob infrastructure based on the above-listed components:

1. Following a **patient's assessment**, the clinician prepares an individual **rehabilitation plan** and an individual **rehabilitation program** (PRIpri) implementation plan for the patient. These contain the description of the rehabilitation objectives and the way to achieve them, i.e. the temporal progression and intensity of the exercises to be performed within the period of rehabilitation activity. The plan is implemented using the clinical dashboard and with the support of a Decision Support System (DSS).
2. The PRIpri implementation plan is transferred to the **delivery settings**, which can be @hospital, @rhab or @Home. The delivery environment has therefore the goal to manage the personalized execution of the exercises. The same plan may be transferred to the app installed on the patient's smartphone, capable of performing personalized virtual coaching functions.
3. The delivery setting, once installed and connected to the **network**, can manage the assisted execution of the entire rehabilitation activity implementation plan, as foreseen by the clinician. The patient performs the exercises connected to an exercise monitor, wearing inertial sensors (IMU), remaining within the range of a 3D camera (Kinect) and interacting or wearing the robotic device. All these devices are connected, mainly wirelessly, to a local-hub and to the virtual environment. At the end of the exercise session, the patient can express the degree of dyspnea perceived by filling in the Borg Scale, used in the clinical setting to evaluate exercise tolerance.
4. The descriptive data of each session performed by the patient are sent, in asynchronous mode and at the end of each session, to a **data lake** in a *private cloud*. Also other session data (signals, outputs of the robotic platform) can be sent to the data lake.
5. The data collected in the data lake are cleaned, organized, and stored in a semantically consistent way in the project **data repository**, that is connected to the clinician dashboard.
6. Using the **clinical dashboard**, the clinician in charge is able to monitor the progress of the rehabilitation activities carried out by the patient by selectively and at different levels of detail, displaying a set of performance indicator parameters. The clinician can modify the PRIpri implementation plan to adapt it to the specific situation encountered. The changes are passed onto the affected components. An auditing system will store any changes to the PRIpri plan.
7. Through software **apps** (e.g., on the smartphone**)**, the patient can be aware of the assigned rehabilitation activity plan and the progress recorded, in a simple and concise way. The app can also perform **virtual coaching** functions, sending reminders and recording additional information, such as the execution of activities.
8. The virtual coaching functions performed by the app are completed possibly by the presence of a **chatbot**, to which it is possible to ask questions in natural language (written or spoken) and have answers and clarifications, of a clinical and technical nature. The chatbot also performs a fast psychological profiling function, to be performed only once: based on the result, the chatbot is able to adapt its responses to the specific characteristics of the patient.

4 Delivery of the robotic intervention

4.1 Robotic platforms prescription and measurements of prescription's outcomes

Starting from the PRI, the clinicians set up the rehabilitative intervention, as well as the robotic platforms (and eventually virtual environment configuration).

The parameters to be configured depends on the rehabilitation task, the specific robotic platform, and the patient's functioning, disability, and health status (that may be described as in Section 3.c with the International Classification of Functioning, Disability and Health (ICF) index).

As an example, parameters that could be configured are: number of repetitions of the tasks/exercise, complexity level, duration of the exercise.

The measurement of the prescription's outcomes has the same peculiarities, being highly dependent on the specific robotic platform and on the specific disease and/or rehabilitation tasks.

In general, the following parameters can describe the outcome: the number of repetitions correctly performed and completed, the time to complete each repetition and the total time of the interventions, the complexity levels of the completed tasks, and the ICF after the interventions. Such measurements could be an indicator of the overall status of the patient after the robotics intervention.

4.2 Virtual environments and exergames

4.2.1 Hardware devices

Virtual (and Augmented) Reality Environments (VR/AR) can be characterized by:

1. the level of immersion, described with respect to the used devices for visualization[1];
2. the coexistence of real and virtual elements[2].

Considering the level of immersion in VR, i.e., the systems where users act in a completely computer-generated environment, the following visualization methods are possible:

a. **(fully) immersive VR systems**, in which the user is completely isolated from the external (real) world. In such systems, the sight of the surrounding environments is hampered by a head-mounted display (HMD). Typically, other senses are hidden, e.g. sounds.
Several commercial[3] immersive VR devices are available: HTC Vive Pro Series[4], or Meta Oculus Quest 2[5]. Such wearable systems are composed of 2 displays for stereoscopic visualization, equipped with lenses and tested for an optimal view at a distance of about 7 meters, a tracking system to compute the 6DOF of the user and update the displayed images, and controllers to interact with VR objects.
Main advantages:
   - The virtual experience is completely detached from the real one, so the users can be present in the virtual situation.
   - Commercial headsets are characterized by good and robust tracking systems, working in various experimental conditions, and affordable costs.
   - Stereoscopic visualization allows a better perception of depth and distances (stereoscopic cues in addition to other depth cues, such as shadows and perspective), this could be important when exergames are used to perform tasks like reaching.

Main issues:
   - Though immersion has been extensively used in the context of gaming, but also for training and teaching, it may not be the optimal solution for rehabilitation purposes and specific populations of users.
   - Literature reports undesired effects after the usage of immersive VR HMDs, e.g. sickness or eye strain.

[1] Slater, M. (2003). A note on presence terminology. Presence connect, 3(3), 1-5.

[2] Milgram, P., Takemura, H., Utsumi, A., & Kishino, F. (1995, December). Augmented reality: A class of displays on the reality-virtuality continuum. In Telemanipulator and telepresence technologies (Vol. 2351, pp. 282-292). Spie

[3] Note that for all wearable device electromagnetic compatibility might be previously assessed.

[4] https://www.vive.com/eu/

[5] https://www.meta.com/it/

b. **semi-immersive VR systems**, in which the user is in front of several "standard" displays arranged in a multi-monitor configuration to extend the field of view. In this configuration, they partially hide the surrounding real environment. Such systems are widely used in simulation scenarios.

   Main advantages:

   - The user is not completely isolated from the real environment, so he/she can see and interact with people (e.g., the therapist or the caregiver).
   - Interaction with physical real objects is possible;
   - The setup requires standard hardware (e.g., monitors) and standard PCs. The tracking of the patient is optional (not necessary to update the visualization)

   Main issues:

   - Usually, visualization is not stereoscopic, thus hampering the correct estimation of the depth of virtual elements. Stereoscopic visualization is possible only with specific systems, e.g. the CAVE.
   - The sense of presence inside the VR environment is negatively affected by low immersivity.

c. **non-immersive VR systems**, in which the user is in front of a standard display (e.g., PC monitor or a smartphone/tablet). Though the level of immersion, thus the sense of presence, cannot be like the ones of immersive VR systems, they are the most common and flexible solutions.

   Main advantages:

   - A large number of off-the-shelves and low-cost devices can be used, no specific hardware is necessary.
   - The tracking of the patient is optional (not necessary to update the visualization)

   Main issues:

   - Stereoscopic visualization is not possible, thus hampering the correct estimation of the depth of virtual elements.
   - The sense of presence is very low.
   - Interaction with virtual elements is achieved with standard input modalities (mouse, touchscreen, joypad) or with vision-based techniques (Leap Motion, RGBD devices, skeleton tracking, see Section 2.d)
   - For these reasons, non-immersive VR is commonly used to implement exergames and serious games rather than interactive virtual environments.

VR systems can be affected by several usability problems, and, in general, interaction with virtual objects may be different with respect to interaction with real ones (e.g., for the lack of haptic and tactile feedback) For this reason, the combination of virtual and real may overcome such issues: real environments can be augmented by inserting virtual elements (Augmented Reality, AR). AR devices can be divided into two main categories:

1. **video see-through (VST)**, where the real environment is captured by a standard camera, and an augmented visualization is obtained by combining the camera feed and a computer-generated image. Typically, VST systems are based on standard smartphones and tablets.
2. **optical see-through (VST)**, where an optical combiner allows the user to see both the real environment and computer-generated images reflected on it. Several commercial devices using this technology exist, e.g. the Microsoft HoloLens[6]. The main issue associated with such devices is the limited field of view, which could negatively affect the rehabilitation outcome. Indeed, a limited field of view could result in wider and not natural heads' movements.

To be effective in the specific rehabilitation context, VR and AR environments must be interactive. Thus, the user must perform specific actions, "playing" inside the environment. The interaction may be implemented with several techniques, such as:

- using controllers, e.g. videogames pad, mouse, keyboard, or the controllers sold with the HMDs;
- tracking the user's movements with non-wearable devices, e.g. cameras, markers, or gloves (see Section 2.d, for a detailed analysis of the motion capture systems).

[6] https://www.microsoft.com/it-it/hololens/hardware

A comparison among the different techniques is out of the scope of this document, several works in the literature address the issue, though no standard for interaction could be defined at the moment[7].

4.2.2 Software

The tasks may be "gamified", so the users are immersed in VR/AR environments and act inside them like in video games (serious games and exergames). In particular, in exergames, the users are asked to perform specific actions, and the level/complexity of the game changes accordingly to their movements. As an example, an exergame could show to the users virtual objects/targets to be reached/grasped. To accomplish the task, the users should perform specific actions, thus completing the rehabilitation tasks.

VR/AR environments, serious games, and exergames are standalone applications running on PCs and/or mobile devices. Game engines, such as Unity 3D or Unreal, are used to design and program them. It is worth noting that Unity3D and Unreal allow multi-platform deployment, so the same software can be distributed onto several hardware platforms. Recently, web-based VR has been possible thanks to WebXR, an API for web content and apps, which are used to support rendering 3D scenes to hardware designed for VR and AR. WebXR makes possible the use of exergames at home, in general, outside laboratory or hospital facilities.

A recent survey analyzes the existing solutions in cognitive assessment and training in terms of adopted visualization and interaction technologies, validation, chosen experimental designs, data collection, and analysis[8]. Despite a general growth in interest and in the number of prototypes developed in the field, standardization of software developments, visualization and interaction devices, and data collection is still missing.

VRPN (Virtual-Reality Peripheral Network) is a device-independent, network-based interface for accessing virtual-reality peripherals in VR applications[9]. VRPN includes drivers for many tracking devices. It also provides an abstraction layer that makes all devices of the same base class look the same; for example, all tracking devices look like they are of the type vrpn_Tracker. This merely means that all trackers produce the same types of reports. At the same time, it is possible for an application that requires access to specialized features of a certain tracking device (for example, telling a certain type of tracker how often to generate reports) to derive a class that communicates with this type of tracker.
Some other middlewares have been developed to give unified access to different resources in VR, most of them focused on motion tracking data exchange (see Section 2.d)

As a goal of Mission 2 in the Fit4MedRob project, we define a general scheme for Virtual Environments and exergames that could be adapted to specific contexts (see Fig.2)

[7] Reski, N., & Alissandrakis, A. (2020). Open data exploration in virtual reality: a comparative study of input technology. *Virtual Reality*, *24*(1), 1-22.

[8] Bassano, C., Chessa, M., & Solari, F. (2022). Visualization and Interaction Technologies in Serious and Exergames for Cognitive Assessment and Training: A Survey on Available Solutions and Their Validation. *IEEE Access*, *10*, 104295-104312.

[9] https://vrpn.github.io/

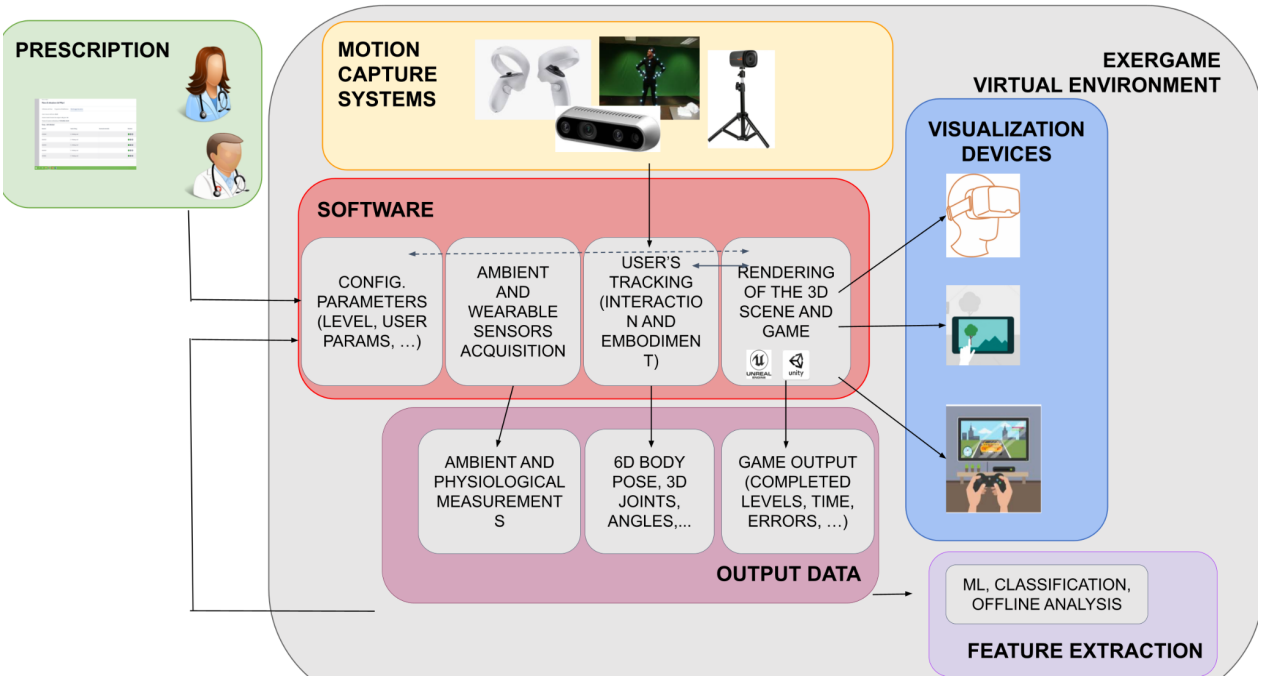


Figure 2: General scheme for Virtual Environments and Exergames

4.3 Sensing systems

Wearable devices are used in rehabilitation to provide biofeedback about biomechanical (measurements of movement, postural control, and force, see also Section 2.d) or physiological body parameters to improve outcomes in people with neurological diseases[10]. Notwithstanding the large use of such measurements, it is not yet clear what the most commonly used sensor configurations are, and it is also not clear which biofeedback components are used for which pathology. Also, acquired data are heterogeneous, and how they are stored, analyzed, and used to evaluate the rehabilitation outcome. Physiological body parameters include:

1. Neuromuscular biofeedback: Electromyography (EMG)

EMG uses surface electrodes to detect a change in skeletal muscle activity, which is then fed back to the user usually by a visual or auditory signal. An EMG signal acquisition system consists of four main stages: (i) signal collection (ii) signal amplification, (iii) signal filtering, and (iv) analog-to-digital converting. Each stage demands specific requirements according to its operational characteristics.

There is a diverse range of commercial systems developed for the monitoring of EMG signals, these may vary according to the number of synchronized channels and the use of these signals.

2. Physiological data and Cardiovascular biofeedback: Heart rate variability (HRV) or respiratory sinus arrhythmia (RSA)

RV refers to the variability in the time between heartbeats. These variations in HR are regulated by the autonomic nervous system. HRV at the frequency of respiratory, which is also termed RSA, refers to the increase in HR with

[10] Bowman T, Gervasoni E, Arienti C, Lazzarini SG, Negrini S, Crea S, Cattaneo D, Carrozza MC. Wearable Devices for Biofeedback Rehabilitation: A Systematic Review and Meta-Analysis to Design Application Rules and Estimate the Effectiveness on Balance and Gait Outcomes in Neurological Diseases. Sensors. 2021; 21(10):3444. https://doi.org/10.3390/s21103444

inspiration and the decrease in HR with expiration. HRV and RSA provide biofeedback on the cardiovascular system, and both terms are used interchangeably in the literature. Although the gold standard for HRV registration is the electrocardiogram (ECG), owing to innovative technological advances, devices with greater portability, lower cost, easier operation, and increased accessibility have recently emerged as new tools for HRV recordings and analysis[11]. The use of these devices, such as chest heart rate monitors (HRM), despite optimizing HRV evaluation, raises issues regarding the reliability and accuracy of the generated data. Although the reproducibility of HRV devices has been evaluated in different situations, whether their performance is equivalent to the gold standard remains controversial.

In addition, physiological data such as heart rate, skin conductance, oxygen saturation, and body temperature acquired by wearable devices are becoming an important source of information for monitoring general health status and psychological stress of a patient. In particular, the use of wearables can provide real-time data on a patient's physiological parameters, allowing clinicians to monitor progress and adjust treatment plans accordingly. Moreover, the integration of deep learning techniques with physiological data acquired by wearables is gaining increasing attention for the detection and monitoring of mental stress and psychophysical state[12]. Deep learning models can be trained on the data stored into a data lake (see section 3 b. Data ingestion and data lake) to detect patterns and identify markers of stress. For example, changes in heart rate variability, skin conductance, and other physiological parameters can be used as indicators of stress levels. By combining wearable devices with deep learning techniques, it is possible to develop personalized stress monitoring systems that can provide timely and accurate feedback to patients, allowing them to manage their stress levels and improve their overall well-being.

In such cases, referring to the overall ICT infrastructure's schema of Figure 1, devices like smartphones act simultaneously as sensing and visualization devices for VR/AR and exergames.

Most previous systems develop processing and visualization in closed-source hardware/software systems. However, the commercial software of these owners often has the opportunity to provide ways to export files for use on other platforms such as Matlab, Excel, Labview among others. The export of these can be done in different types of files: Binary, CSV or Matlab files. Currently, there are also systems based on platforms like Arduino or Raspberry Pi, whose hardware, software, and mechanical design files are available online offering a solution to the high costs and slow pace of innovation of medical devices, in such cases standardization of the acquired data is missing.

In Section 3, we describe the reference data model for representing the data collected during the monitoring and the delivery of the intervention (potential solutions are LOINC and SNOMED).

4.4 Wearable sensors and motion capture systems

Motion Capture technology (MoCap) is the process of recording the movement of people or objects and transferring the data to a computer to animate digital characters.

In our context, the movements of the patients are captured for two purposes:

- to interact with the VEs, the exergames, and the robots. In this situation, the 6DOF of the patient's limbs are tracked to align the patient, the robot, and the VEs reference frames, so that they could be represented in the

[11] Oliveira Júnior FA, Pereira RA, Silva AS, Brito Alves JL, Costa-Silva JH, Braga VA, Balarini CM. Different acquisition systems for heart rate variability analysis may lead to diverse outcomes. Braz J Med Biol Res. 2022 Feb 4;55:e11720. doi: 10.1590/1414-431X2021e11720. PMID: 35137854; PMCID: PMC8852161.

[12] Gedam, Shruti, and Sanchita Paul. "A review on mental stress detection using wearable sensors and machine learning techniques." *IEEE Access* 9 (2021): 84045-84066.

same spatial location. In this context, tracking of the users is also necessary to represent them (or parts of their bodies, e.g. the hands) inside the VE, e.g. to create a self-avatar.
- to assess the quality of the movements, thus the intervention outcome. In this case, the 6DOF of the patients' joints is further analyzed to compute specific qualities to assess the rehabilitation outcome and/or the patient's physical state.

The same tracking technique (see Hardware) could be used for both purposes. However, it is worth noting that the requirements in terms of robustness and accuracy of the tracking could be different. In the following, we will describe the Hardware and Software in relation to the MoCap technology.

4.4.1 Hardware

Motion capture systems are based on different tracking technologies. Measurement systems used in tracking can employ a variety of physical phenomena and arrangement options. These choices determine which coordinate systems are being measured and affect which temporal and spatial properties the tracking has.

Moreover, tracking systems could be divided into:
- wearable devices, which include all the systems that require sensors or sources to be worn by the users/patients
- non-wearable devices, which rely on sensors and sources placed in the environment and not physically attached to the users

Contemporary motion capture users can utilize one of the following techniques:

1. Mechanical tracking

In this case, the end-effector of an articulated arm with several limbs and joints is tracked. This requires knowledge of the extent of every limb and measurement of the angles at every joint. Joints can have one, two, or three degrees of freedom in orientation, measured using rotary encoders or potentiometers. A kinematic chain can be set up to determine the position and orientation of the end-effector. Such setups can also provide force feedback.

2. Inertial

This technique records movement through inertial measurement units (IMUs) such as gyroscopes, magnetometers, and accelerometers, containing sensors to measure rotational rates. An electronic gyroscope is a device for measuring rotational velocity. With numerical integration, the orientation can be computed. Three orthogonal gyroscopes are usually combined in a micro-electromechanical system (MEMS) to deliver a full 3DOF orientation measurement. Inertial sensors are sourceless but provide only relative measurements, and so are rarely used alone. They have high update rates, up to 1000 Hz. However, the required integration makes them susceptible to accumulated drift. Inertial position measurement is obtained with linear accelerometers. Also built using a MEMS approach, this device allows sourceless estimation of accelerations. After subtracting the effects of gravity and integrating twice numerically, the position can be computed from the acceleration measurements. Examples of systems that use IMU-based tracking are Xsens MVN, Perception Neuron, or Rokoko Smartsuit.

3. Optical (Passive)

Retroreflective markers attached to bodies or objects reflect light generated near the camera lens to calculate the position of the markers within a three-dimensional space and are recorded. The camera's sensitivity to capturing the light can be adjusted so that only the reflective markers are sampled, ignoring skin and other materials. The motion capture system estimates the 3D position of the centroid of each marker from the 2D view captured by individual cameras through triangulation. Optical passive mocap systems are widely used in biomechanics laboratories. The main advantages are the robustness and accuracy of the measurements, the main drawbacks are the costs and the necessity of precise and complex calibration, usually performed by expert technicians.

4. Optical (Active)

Using the same technique as above except the markers emit light instead of reflecting it and require a power source. Advantages and drawbacks are similar to passive solutions.

Optical solutions that are based on passive or active (infrared) markers include OptiTrack Motive:Body, Qualisys Track Manager, ART-Human, Vicon Tracker, and Motion Analysis Cortex.

5. Marker-less

Instead of relying on markers, this method uses standard (RGB) and/or depth-sensitive cameras (RGBD) and specialized software to track and record movement. The main advantage is that no sensors/markers are required on the users/patients, and off-the-shelves hardware (RGB cameras or RGBS ones) can be used. The main issue is the accuracy of the technique compared to the alternatives.
The most important RGBD devices are relative to the family of Microsoft Kinect. In the last decade, Microsoft has released 2 devices (i.e., Kinect v1 and Kinect v2) for gaming, which have been exploited by scientific researchers to develop marker-less MoCap solutions in several research fields. In 2020, Microsoft released the Kinect Azure which has been totally designed as RGBD device for research and development. Other commercial RGBD devices available on the market are the Interl realSense Family. A comparison of the different devices is out of the scope of this document. All the devices provide a sequence of standard RGB images and a sequence of depth images (i.e. images in which the information associated with each pixel is the distance in m with respect to the camera). Post-processing on this raw data is usually done by the associated SDKs to provide skeleton information (see next Section)
Recent use of Convolutional Neural Network (CNN) allowed the research community to estimate 3D pose information from standard RGB images (e.g., DeepLabCut[13], OpenPose[14]).

4.4.2 Software

In the previous section, we briefly analyzed the different motion capture systems (MoCap), which cover a wide variety of different technology, quality, and price range. Such systems are used for character animation, dancing, or gaming, but also for human body analysis and for placing the user replica (avatar). Almost every MoCap streams out different protocols and tracking data. They vary in terms of scale and offset, and their skeletal data differs in rotational offsets between joints and in the overall number of bones. Due to this circumstance, techniques are not effortlessly interchangeable. Usually, software that makes use of a technique is rigidly bound to it, and a change to another system can be a complex procedure. In Section 2.b, we already discussed VRPN as a network-based interface to allow communication among devices and VR systems.
Focusing on MoCap, the first official international standard for humanoid animation is the H-Anim, created within the scope of the Extensible 3D (X3D) standard and is a successor of the Virtual Reality Modeling Language (VRML). H-Anim has been described in ISO/IEC 19774-2:2019[15] and is one of the only efforts yet to create an official open standard for humanoid avatar motion and data exchange.

COLLADA and FBX interchange file formats for 3D applications and are widely used today. They are used both for saving body tracking data and for describing 3D geometry. The Biovision Hierarchy (BVH) file format was developed exclusively for handling skeletal motion data and is therefore simpler in structure. It is supported by many body tracking applications, and because of its simplicity and less overhead compared to other file formats, it is often used for real-time transmission of humanoid motion data.

Other software and data formats have been developed by hardware producers. Some work only with the corresponding device, while others are cross-platform systems. As an example, in 2010 Microsoft started shipping the Kinect (see previous section) and the associated SDK. Microsoft Kinect allowed the body tracking community to grow substantially since it was an affordable and capable sensor. During this time, PrimeSense made OpenNI[16] and NITE[17] publicly available. OpenNI offered low-level access to the Microsoft Kinect and other PrimeSense sensors, while NITE was a middleware that enables the user to perform higher-level operations such as body tracking and gesture recognition. Recently, PrimeSense stopped the distribution of OpenNI and NITE.

[13] https://github.com/DeepLabCut
[14] https://github.com/CMU-Perceptual-Computing-Lab/openpose
[15] https://www.iso.org/standard/64791.html
[16] https://structure.io/openni
[17] https://wiki.debian.org/PrimeSenseNite

After that, many proprietary MoCaps such as iPi Mocap Studio, Brekel Body, or Nuitrack[18], which are based on RGB-D streams, have been developed.

OpenPose (one of the main markerless MoCap systems relying on RGB cameras) streams skeleton data in standard JSON formats, compatible with OpenCV libraries. A Unity3D plugin is available.

It is also worth mentioning that the main MoCap systems now provide direct integration to VR development software (e.g. Unity3D, see Section 2.b)[19] [20].

Recently, the research community addressed again the problem of having a standard for skeletal tracking. As an example, the middleware solution MotionHub[21] can receive and process data of different MoCap technologies. It converts the spatial as well as the skeletal tracking data into a standardized format in real-time and streams it to a client (e.g. a game engine). That way, MotionHub ensures that a client always receives the same skeletal-data structure, irrespective of the used MoCap system.

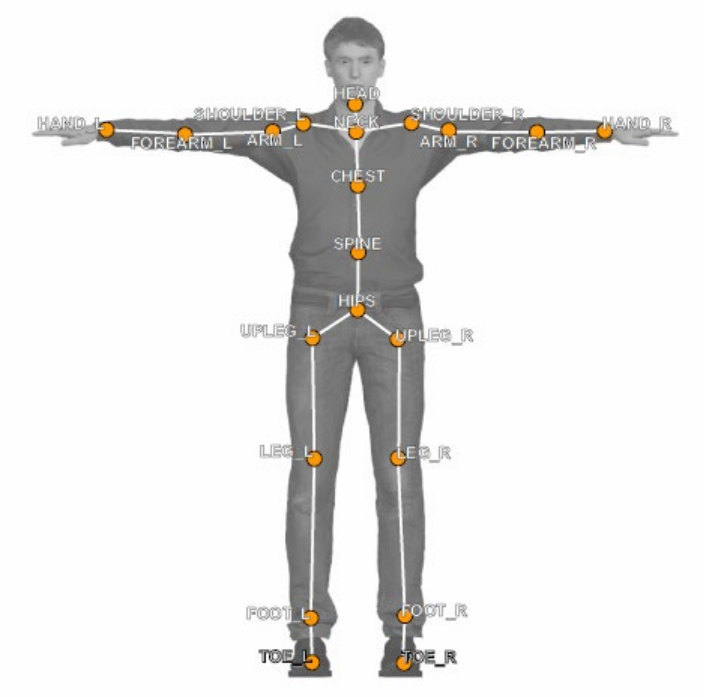


Figure 3: A possible unified skeleton structure (figure from Motion -Hub)

4.5 Ambient sensors

The concept of sensing the ambient typically arises in the context of elderly assistance, e.g., in smart homes, in which various types of sensors/devices are integrated into everyday objects. In this context, the techniques could be extended to monitor the patients when the rehabilitation intervention is delivered at home or in a protected care service.

[18] https://nuitrack.com/

[19] https://www.qualisys.com/integrations/unity/

[20] https://optitrack.com/software/unity/

[21] https://github.com/Mirevi/MotionHub

Infrastructure in the smart home is connected by network technologies for gathering contextual information such as vital signs and behavioral information of the elderly via the sensors. The most common approaches for patient monitoring are based on machine vision. In this case, stereo or multi-camera configurations (using standard RGB devices), or RGBD devices (e.g., Microsoft Kinect or Intel RealSense) could be considered.

However, other sensors (e.g., motion, radar, object pressure,audio, and floor vibration sensors) are also used for health and behavior monitoring.

Among the many challenges that are encountered in implementing assistive technology in the home, one key challenge is related to the continuous observation of the vital signs and behaviors of elderly subjects through non-wearable ambient sensors. The challenge is related to important factors such as durability, acceptability, communication, and power requirements of the sensors installed in the sensorized environments.

In the literature, a number of different setups with different combinations of sensors are described. Ambient sensors are basically installed in several places in a smart setting to acquire user data and send it to a base machine (typically using wireless communication) for further processing.

Smart healthcare projects often try to utilize consumer-targeted technology. The technology includes a set of sensors and devices for monitoring the users and predicting problems related to emergency cases. At the same time, the volumes of sensor data rapidly grow which makes the data more complicated and difficult to manage. Moreover, there is a lack of suitable outcomes to validate the installation, management, and delivery of technological solutions to meet specific needs. Insufficient experience with smart healthcare initiatives has demonstrated that pilot projects do not always lead to an extensive scale of technology applications.

Finally, privacy risks and possible intrusion must be considered. Acceptance of sensors such as video cameras may be challenging, as cameras can easily be perceived as intrusive by patients (and therapists). It is also worth noting that acceptability is culture-dependent and will differ from one society to another.

4.6 The Local Hub for data processing and communication

The Local Hub is the software that:

- reads the inputs (prescription for the patients, VR/AR parameters configuration/robots configurations)
- initialize the VR/AR with the right parameters
- update the VR/AR accordingly to: users actions/ users movements (detected like in 2.d)/ users status (detected in 2.c) / environment status (detected in 2/e)
- record and store output data: users actions (task completed, objects grasped,...)/ users movements (6DOF pose of k users' joints at given fps)/ users status (n values at given fps)/ environment status (m values at given fps)
- transmit stored data to data lake

In addition: RGB and RGBD video recorded are stored for a post analysis, motion quality features, and behavioral parameters can be extracted from raw data.
The local hub also contains the Computer Vision, image Processing, Machine Learning techniques to extract higher level features.

5 Monitoring and assessment of the robotic intervention

5.1 The monitoring communication platform

The main purpose of a monitoring communication platform is that of facilitating communication between the robotic system and the human operator or medical personnel, or even the subject at home. The platform allows medical personnel to monitor and assess the progress of the robotic intervention remotely, either in real-time or off-line.

In any case, the main feature to provide for such a platform is the presence of a **broadband internet connection**: this is a crucial aspect to provide the necessary conditions to ensure secure, reliable and fast data transfer between all the involved actors (clinicians, subjects, care providers) in the different contexts where they operate (@hospital, @rhab or @Home).

It typically includes various components such as sensors, software, and hardware. The sensors collect data related to the robotic intervention; for instance, in cases where exoskeletons or robotic arms are used, sensors measure the position of the robot, the force exerted by the robot, beside general physiological data of the patient (please, refer also to previous Section 2). This data is then transmitted to the software component of the platform, which analyzes and interprets the data. The platform, depending on its deployment site, might also include dedicated communication channels, which enable medical personnel to communicate with the robotic system operator and the patient during the intervention. Such channels are based on communication protocols, such as User Datagram Protocol (UDP) or Transmission Control Protocol (TCP) for real-time communication, or H.323 or SIP (Session Initiation Protocol) for video and audio communication between the medical personnel and the patient, or Application Programming Interfaces (APIs) such as REST (Representational State Transfer) or WebSocket for communication between the user interface software and the data analysis software.
A further comment to connect homes (and in general, non-clinical environments) and hospitals for rehabilitation: several communication infrastructures can be clearly implemented, depending on the specific needs and requirements of the healthcare system. Among the possible available options, we cite Virtual Private Network (VPN), that can be used to connect the home and hospital networks securely over the internet, thus allowing medical personnel to remotely access the communication platform and monitor the patient's rehabilitation progress; cloud-based communication platform, that can also be used to securely store data in the cloud; telehealth software, that can be used to connect the patient and medical personnel remotely (e.g., video conferencing, messaging, and data-sharing features, allowing medical personnel to monitor the patient's rehabilitation progress and provide guidance and support).

The careful design, implementation and deployment of the communication platform has clearly several benefits and advantages. For instance, it allows medical personnel to monitor and assess the progress of the intervention remotely, which can be particularly useful in situations where the patient is in a remote location or where travel is difficult. Additionally, the platform can provide real-time feedback to the robotic system operator, which can help to optimize the intervention and improve patient outcomes.

Finally, a crucial aspect to carefully handle, and that is connected to any kind of communication infrastructure, is the care of security and privacy aspects, both addressed in next Section 5. In particular, security in communication constitutes a major difference between implementing a communication platform service @Home versus in an hospital or a dedicated care service is the security of the communication itself.

5.2 Data ingestion and data lake

In order to finalize the onboarding and data integration process, it is necessary to define the set of rules for collection, metadata creation, classification, discovery, lineage, access, and data usage, as the aim of the platform is not only defining a collector but also ensure the possibility of analyzing data, enabling the evaluation of quality, representativeness, significance, etc.
In these terms, the platform must be capable of uploading data in various formats (images, videos, signals, medical records) according to market reference standards (HL7 and FHIR), both in batch and stream modes.

The source integration process, managed by a data integration layer, will follow the first phase of data collection, which specifically consists of acquiring data from different sources, including the robotic system, electronic health records (EHR), the PNT platform, medical devices, electronic health apps, wearable sensors, and other IoT devices.
The data collection phase must also be supported by data profiling tools (such as frequency analysis or cardinality analysis), with the primary objective of analyzing incoming data and identifying any patterns or anomalies contained therein. Data profiling activities, in this sense, are complementary to data quality assessment activities immediately following them, such as data completeness checks or source consistency checks, as both are aimed at ensuring the highest possible quality standard for the dataset being processed.
Once the data collection phase is complete, including the data profiling and data quality assessment operations described above, the data must be sent to the data cleansing service for cleaning and rectification. Subsequently, they must then be subjected to data deduplication techniques (such as checksum comparison or the use of hashing systems), through which any duplicate records or redundant information contained therein are removed. In this way, it should be possible to improve data management efficiency and reduce the storage space required for their archiving.
After completing the data deduplication step, the subsequent phase, called data integration and handled by the dedicated structural layer, must integrate and harmonize data from different acquisition sources using matching and fusion techniques, generating a more complete and cohesive dataset. The next step must involve the use of data mapping and data transformation techniques to standardize the structure of the acquired data in the shared FHIR standard. In practice, these operations must be implemented through an appropriate "FHIR-transformation engine" capable of converting tabular (and non-tabular) data into the FHIR resource format, making them manageable by the interoperability layer.
The processed and normalized data will be stored in a data foundation layer within a data lake that constitutes the platform's unique repository, which serves as the basis for all the next operations.
Data will be accessible by clinicians or multidisciplinary boards within the hospital in order to let them discuss the case and propose new PRipri. Data accessibility should be granted either during or after the completion of the rehabilitation session. To this end, data will be visualized in the dashboard privileging human readability and interpretability.

5.3 Data model definition and data harmonization

Several types of data are managed and collected by the elements of the three pillars of the Fit4MedRob architecture, namely prescription, delivery and monitoring of the robotic interventions.

The data models used by supporting the three phases can be categorized into a number of classes, including: i) a reference data model to represent the patient's status; ii) a data model to represent the results of the rehabilitation intervention; iii) a data model to represent the prescription of the rehabilitation intervention; iv) a data model to represent the data collected during the monitoring of the intervention.

The first three classes can be modeled at a high level by resorting to the same conceptual model. In fact, in this case, while the low granularity representation of the data elements may be left to the specific implementations, a reference data model may be provided to bind to a unique approach to describe the diseases, the patient's functioning, disability and health status, and thus the rehabilitation robotics interventions. To this end, the proposal of this deliverable is to resort to the International Classification of Functioning, Disability and Health (ICF).

The concept of functioning has recently been recognized as the third health indicator in medicine; this statement is aligned with recent epidemiologic surveys showing non-stop population aging and the progressive growth of chronic conditions. Now more than ever, an urgent need for advances in management improvements in the field of rehabilitation is claimed. The introduction of the ICF classification system represents a landmark event for rehabilitation. The ICF, when eventually used in combination with the ICD, is on its way to provide a holistic model to develop comprehensive functioning profiles and patient-oriented interventions.

ICF sets represents a methodological standard for the functioning (organs and diagnosis involved), disability (symptoms and signs), social participation restriction, and environment interaction assessment intending to provide interdisciplinary and patient-oriented interventions. The ICF describes disability as the result of interactions between the individual, society, and the environment in terms of both barriers and facilitators.

An individual's health and health-related states can be recorded by selecting the appropriate category code or codes and adding qualifiers, which specify the extent or severity of functioning or disability in that category, or the degree to which an environmental factor represents a facilitator or barrier. Patients present a huge variability of disability and participation needs: ICF classification could be informative about the effect of chronic and acute diseases on individuals and help to develop a more comprehensive treatment plan in the rehabilitation setting to merge and not to replace well noted specific outcomes.

The ICF classification helps to consider specific aspects of functioning, context and subject, which insist on the clinical and functional diagnosis, guaranteeing a broader vision of the management and of the treatment path in general.

In summary, ICF:

- provides a scientific basis for understanding and studying health as an interaction between the individual and the context;
- constitutes a common language for describing health and related conditions, with the aim of improving communication between health professionals, researchers, planners, public administrators and the population, including people with disabilities;
- allows the comparison between data collected in different countries, healthcare disciplines, services and moments;
- provides a systematic way to encode information in health information systems.

The ICF focuses on three components, which underscores the importance of the interplay and influence of both internal and external factors to each individual's health status (see Figure 4):

- Body Functions and Structures. Examples: b28010 Pain in head and neck, s720 Structure of shoulder region;
- Activities and Participation (at individual and societal levels). Examples: d230 Carrying out daily routine; d920 Recreation and leisure;
- Personal and Environmental Factors (at a contextual level). Examples: e115 Products and technology for personal use in daily living; e355 Health professionals.

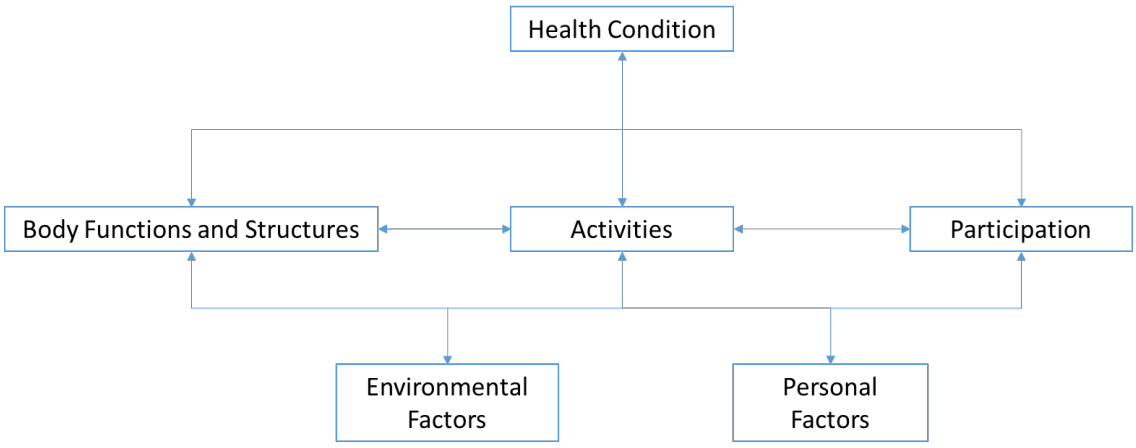


Figure 4: The ICF Model

Each component is divided into a hierarchy with an additional digit added to the classification code for each subsequent layer in the hierarchy. The hierarchy is as follows:

- Component e.g. Activities and participation
- Chapter e.g. Mobility (Chapter 4)
- Block e.g. Walking and Moving (d450-d469)
- Two-level Category e.g. Moving around in different locations (d460)

- Three-level Category e.g. Moving around within the home (d4600)

To record the extent of the problem for each identified impairment, activity limitation and participation restriction, it is possible to use a generic qualifier scale. Environmental factors can also be qualified as either barriers or facilitators.

<table>
<tr><td>Parts</td><td colspan="2">Functioning and Disability</td><td colspan="2">Contextual factors</td></tr>
<tr><td>Components</td><td>Body Functions and Structures</td><td>Activities and Participation</td><td>Environmental factors</td><td>Personal factors</td></tr>
<tr><td>Definitions</td><td>Body Functions are the physiological functions of the body systems
Body Structures are anatomical parts of the body</td><td>Activity is the execution of a task or action by the individual
Participation is the individual’s involvement in his/her life situation</td><td>Environmental factors make up the physical, social and attitudinal environment in which people live and conduct their lives</td><td>Personal factors are the particular background of an individual’s life and living, and comprise features of the individual that are not part of the health condition or health states</td></tr>
<tr><td rowspan="2">Positive aspects</td><td>Functional and structural integrity</td><td>Activity and Participations</td><td rowspan="2">Facilitators</td><td rowspan="2">Not applicable</td></tr>
<tr><td colspan="2">Functioning</td></tr>
<tr><td rowspan="2">Negative aspects</td><td>Impairments</td><td>Activity limitations
Participation restrictions</td><td rowspan="2">Barries</td><td rowspan="2">Not applicable</td></tr>
<tr><td colspan="2">Disability</td></tr>
</table>

As Fit4MedRob is concerned, there are several ICF Core Sets that can be used for our purpose. The ICF Core Sets were developed as a practical tool to facilitate the ICF use in clinical practice. Core sets for twelve chronic diseases were initially developed because of their prevalence and the significant impact on function they can cause. Additional Core Sets have been subsequently developed for other conditions and populations. Those of interest include: Stroke, Amputation, Cerebral Palsy, Hand conditions, Multiple Sclerosis, Rehabilitation, Spinal Cord injury and Traumatic Brain Injury.

For example, a data collection system based on ICF core sets about patients' with stroke can be performed with a form as the one shown in the following:

PATIENT INFORMATION

BODY FUNCTIONS Physiological functions of body systems (including psychological functions) ***How much impairment does the person have in ...***		No impairment	Mild impairment	Moderate impairment	Severe impairment	Complete impairment	Not specified	Not applicable
		0	**1**	**2**	**3**	**4**	**8**	**9**
b110	**Consciousness functions**	☐	☐	☐	☐	☐	☐	☐
	General mental functions of the state of awareness and alertness, including the clarity and continuity of the wakeful state. Inclusions: functions of the state, continuity and quality of consciousness; loss of consciousness, coma, vegetative states, fugues, trance states, possession states, drug-induced altered consciousness, delirium, stupor Exclusions: orientation functions (b114); energy and drive functions (b130); sleep functions (b134) **Sources of information:** ☐ Case history ☐ Patient reported questionnaire ☐ Clinical examination ☐ Technical investigation **Description of the problem:**							
		0	**1**	**2**	**3**	**4**	**8**	**9**
b114	**Orientation functions**	☐	☐	☐	☐	☐	☐	☐
	General mental functions of knowing and ascertaining one's relation to self, to others, to time and to one's surroundings. Inclusions: functions of orientation to time, place and person; orientation to self and others; disorientation to time, place and person Exclusions: consciousness functions (b110); attention functions (b140); memory functions (b144) **Sources of information:** ☐ Case history ☐ Patient reported questionnaire ☐ Clinical examination ☐ Technical investigation **Description of the problem:**							

BODY STRUCTURES

Anatomical parts of the body such as organs, limbs and their components

How much impairment does the person have in ...

			No impairment	Mild impairment	Moderate impairment	Severe impairment	Complete impairment	Not specified	Not applicable
			0	1	2	3	4	8	9
s110	**Structure of brain**	**Extent**	☐	☐	☐	☐	☐	☐	☐

	0	1	2	3	4	5	6	7	8	9
Nature*	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Location**	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Sources of information:

☐ Case history ☐ Patient reported questionnaire ☐ Clinical examination ☐ Technical investigation

Description of the problem:

			0	1	2	3	4	8	9
s730	**Structure of upper extremity**	**Extent**	☐	☐	☐	☐	☐	☐	☐

	0	1	2	3	4	5	6	7	8	9
Nature*	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Location**	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

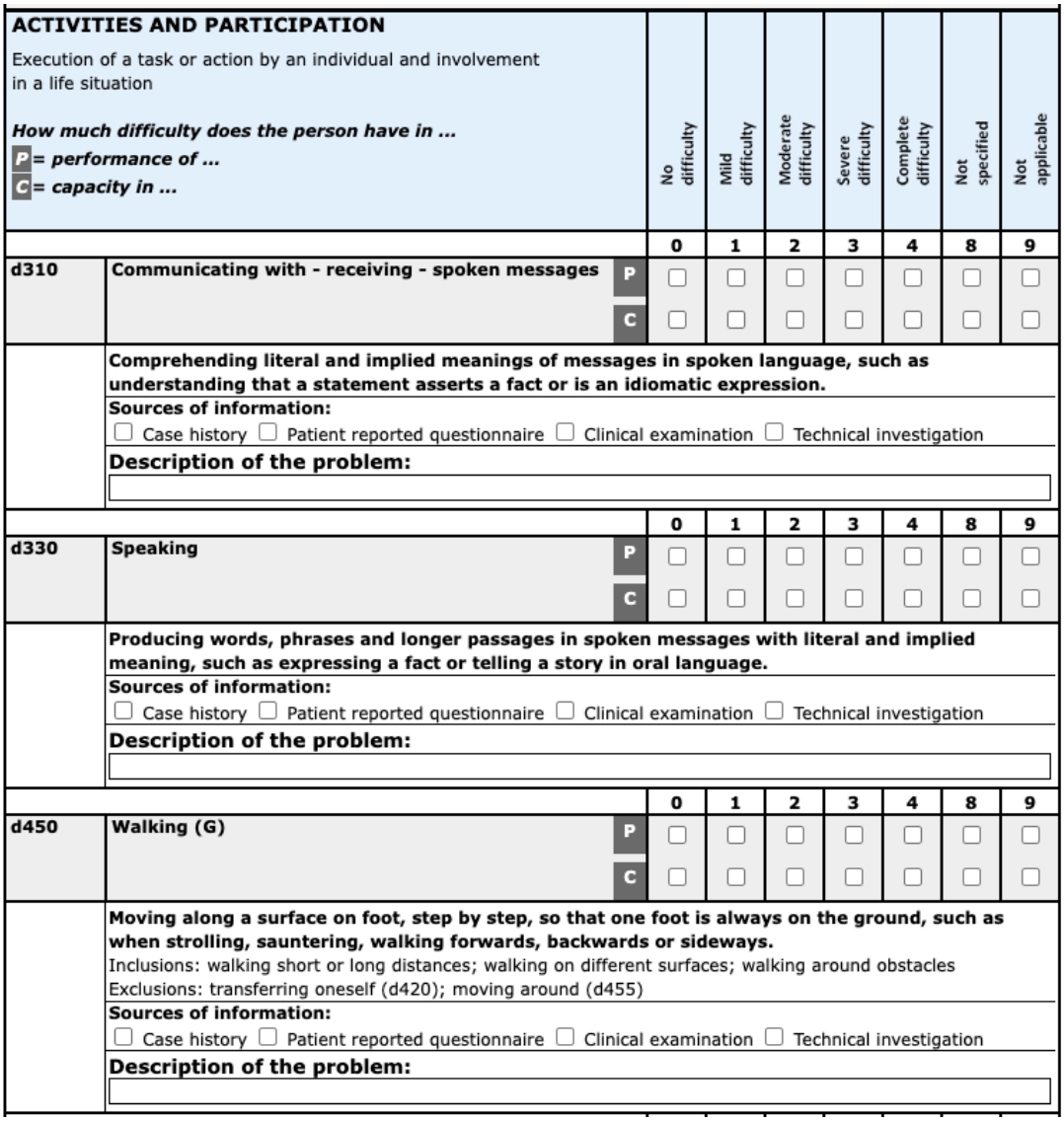
Sources of information:

☐ Case history ☐ Patient reported questionnaire ☐ Clinical examination ☐ Technical investigation

Description of the problem:

* 0 = no change in structure, 1 = total absence, 2 = partial absence, 3 = additional part, 4 = aberrant dimension, 5 = discontinuity, 6 = deviating position, 7 = qualitative changes in structure, 8 = not specified, 9 = not applicable

** 0 = more than one region, 1 = right, 2 = left, 3 = both sides, 4 = front, 5 = back, 6 = proximal, 7 = distal, 8 = not specified, 9 = not applicable



ACTIVITIES AND PARTICIPATION

Execution of a task or action by an individual and involvement in a life situation

How much difficulty does the person have in ...
P = performance of ...
C = capacity in ...

			No difficulty	Mild difficulty	Moderate difficulty	Severe difficulty	Complete difficulty	Not specified	Not applicable
			0	1	2	3	4	8	9
d310	Communicating with - receiving - spoken messages	P	☐	☐	☐	☐	☐	☐	☐
		C	☐	☐	☐	☐	☐	☐	☐
	Comprehending literal and implied meanings of messages in spoken language, such as understanding that a statement asserts a fact or is an idiomatic expression. **Sources of information:** ☐ Case history ☐ Patient reported questionnaire ☐ Clinical examination ☐ Technical investigation **Description of the problem:**								
			0	1	2	3	4	8	9
d330	Speaking	P	☐	☐	☐	☐	☐	☐	☐
		C	☐	☐	☐	☐	☐	☐	☐
	Producing words, phrases and longer passages in spoken messages with literal and implied meaning, such as expressing a fact or telling a story in oral language. **Sources of information:** ☐ Case history ☐ Patient reported questionnaire ☐ Clinical examination ☐ Technical investigation **Description of the problem:**								
			0	1	2	3	4	8	9
d450	Walking (G)	P	☐	☐	☐	☐	☐	☐	☐
		C	☐	☐	☐	☐	☐	☐	☐
	Moving along a surface on foot, step by step, so that one foot is always on the ground, such as when strolling, sauntering, walking forwards, backwards or sideways. Inclusions: walking short or long distances; walking on different surfaces; walking around obstacles Exclusions: transferring oneself (d420); moving around (d455) **Sources of information:** ☐ Case history ☐ Patient reported questionnaire ☐ Clinical examination ☐ Technical investigation **Description of the problem:**								

Figure 5: Patient's information in ICF

Finally, the reference data model for representing the data collected during the monitoring of the intervention, can be built on the basis of the coding and terminology systems defined for measurements and biosignals. Two potential solutions are LOINC and SNOMED.

The **Logical Observation Identifiers Names and Codes** (LOINC) system is the world's most widely used terminology standard for health measurements, observations, and documents. Codes are defined for a variety of biosignals, and can be associated with a variety of clinical relevant parameters collected by the architecture. LOINC enables the exchange and aggregation of clinical results for care delivery, outcomes management, and research.
LOINC codes are used widely in the context of message exchange between a Clinical **Laboratory Information Management Systems** (LIMS) and **Electronic Health Record Systems** (EHR). In this way, LOINC codes provide universal identifiers that allow the exchange of clinical data between heterogeneous computing environments. Such identifiers (names and codes) can be used in the context of observation exchanges between information systems that use syntax standards such as HL7, CEN TC251, ISO TC215, and DICOM. Specifically, the identifier can be used as the coded value for an observation in any other standard that uses the observation/observation value paradigm, whether messages, documents, application programming interface (API), etc.

LOINC focuses on description of the analysis. A specific structure for result values is not part of the standard. LOINC codes only specify the measured property (e.g. mass concentration) but not the reported unit (e.g. mg/dL). To this end, the Unified Code for Units of Measure (UCUM) system is recommended to express units together with LOINC codes. LOINC has codes to describe sample quality (e.g. 20393-5 Sample hemolyzed) but lacks a mechanism to encode interpretive comments or measurement uncertainty.

SNOMED CT (Systematized Nomenclature of Medicine - Clinical Terms) is a standardized, multilingual vocabulary of clinical terminology used for the electronic exchange of clinical health information. It is owned and administered by SNOMED International, a not-for-profit organization.
SNOMED CT currently contains more than 300,000 medical concepts, divided into hierarchies such as body structure, clinical findings, geographic location and pharmaceutical/biological product. Each concept is represented by an individual number (Concept Unique Identifier - CUI) and several concepts can be used simultaneously to describe a complex condition.

SNOMED and LOINC concepts can be consistently mapped, so that our reference architecture can comply with both of them, although LOINC is able to map the data at a higher level of detail, having codes, for example, that describes the components of the PQRST waves, including QT intervals, and R-R intervals.

It is worthwhile mentioning that not all signals collected might have a straightforward mapping into the concepts. In that case, an ad-hoc coding system will be implemented by project's partners.

In summary, the overall proposal for data models and data harmonization is reported below, mapped onto the proposed reference architecture.

RedCap
Research Electronic Data Capture (REDCap[22]) is a web-based application developed by Vanderbilt University to capture data for clinical research and create databases and projects. It is Health Insurance Portability and Accountability Act (HIPAA)–compliant, highly secure, and intuitive to use. The databases use instruments such as surveys and forms as research capture tools. Projects are self-sufficient and secure databases that can be used for normal data entry or for surveys across multiple distinct time points. They are workflow-based and focus on collecting data and exporting it to statistical programs and other data analysis software. REDCap is designed to provide a secure environment so that research teams can collect and store highly sensitive information. Many medical libraries have started using REDCap for assessment and capturing of data for projects.

REDCap is used to collect data for IRB studies, capture data from electronic health records, create surveys, collect data for quality improvement studies, create pre- and post-surveys for library instruction, and track library use statistics. Also, Lyon, Garcia-Milan, Norton, and Tennant wrote an article describing how they use REDCap to create data in records for current searches[23].
There are several advantages in organizing data though RedCap: there is no need to know programming to set up a database or project in REDCap. However, users who know how to program can use an application programming interface (API) for developing mobile applications and more dynamic data import and export. RedCAP is more secure than Microsoft Excel or Microsoft Access and can be accessed from any device with an Internet connection and web browser. It is also HIPAA-compliant; fields in REDCap can be marked as identifiable; and the user has the option of de-identifying their data during export. REDCap also offers daily backups, basic support, and an audit trail feature for even more security. REDCap provides easy exports so users are in control of their data.

[22] https://projectredcap.org/about/

[23] 1. Lyon JA, Garcia-Milian R, Norton HF, Tennant MR. The use of Research Electronic Data Capture (REDCap) software to create a database of librarian-mediated literature search. MedRef Serv Q. 2014;33(3):241–52. doi: 10.1080/02763869.2014.925379.

However, RedCap is not the only tool available for clinical research data capture: in fact, there are more than eighty competitors[24], especially in this arena of clinical trials. Two examples are IBM Clinical Development and Videos. Both are comparable to REDCap, because they offer features such as data entry from anywhere and data capture design capabilities. REDCap also has some features in common with survey creation tools such as SurveyMonkey and Qualtrics. However, according to Bas De Veer, the lead REDCap administrator at the UW Institute of Translational Health Sciences, REDCap differs from these tools in that it is built by clinical researchers specifically for clinical research. It is a clinical research database first and a survey tool second.

From a technical perspective, RedCap infrastructure requires[25]:

1. Web server (e.g. Microsoft IIS or Apache) with PHP 7.2.5 and higher (including support for PHP 8);
2. Database server with MySQL 5.5.5+ or MariaDB 5.5.5+ a. MySQL client – required for installation/upgrades (e.g. phpMyAdmin, MySQL Workbench);
3. SMTP email server – In order to send emails from REDCap, an SMTP server must be configured with PHP on the web server. It can be installed on the same web server or on a separate server (preferred), such as an existing institutional SMTP server, if available;
4. File server (optional) – Depending on the infrastructure and setup, it might be useful to employ a separate server solely for files uploaded/stored in REDCap. If the web server is accessible to the web, it is also recommended to have a separate file server located behind a firewall that communicates securely to REDCap using WebDAV protocol (SSL supported).

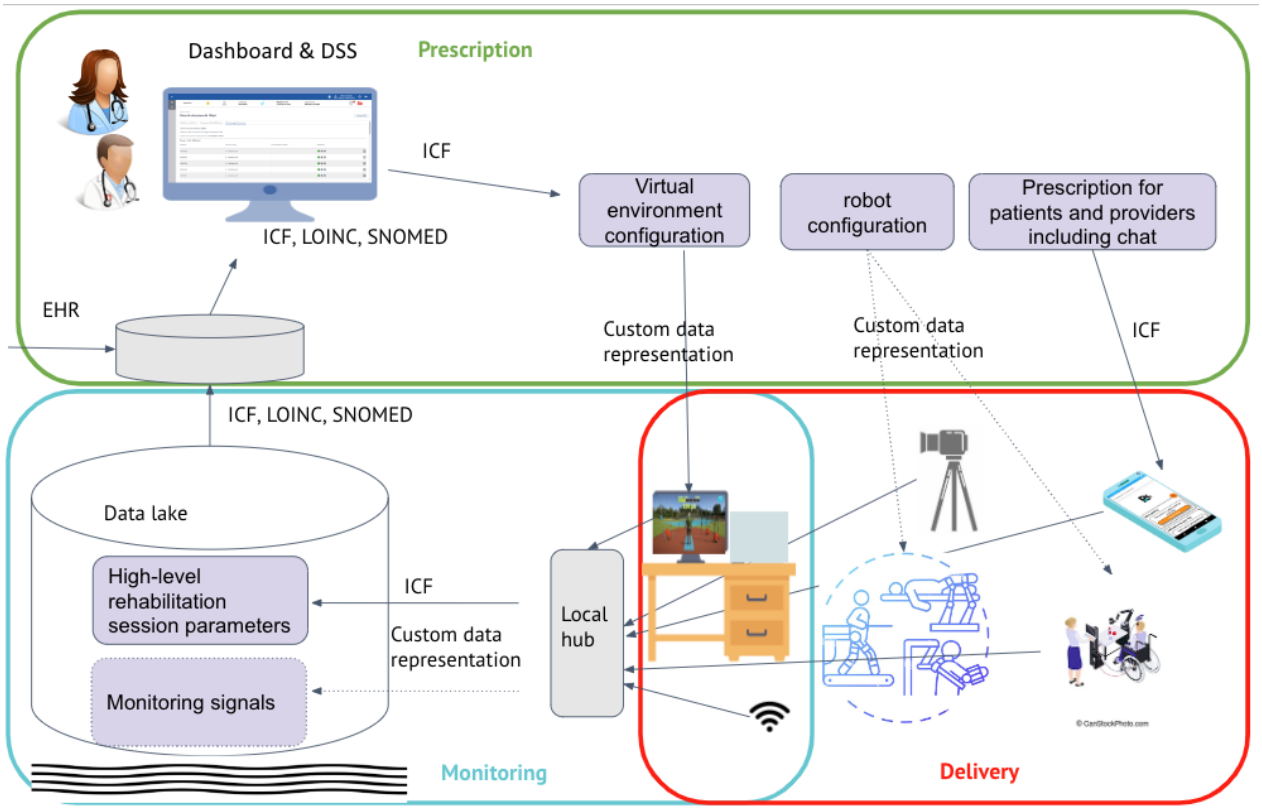


Figure 6: The overall Fit4MedRob schema

[24] Capterra. Top electronic data capture software products. Capterra; 2017.

[25] https://projectredcap.org/wp-content/resources/REDCapTechnicalOverview.pdf

5.4 Data repository and Integration with the Electronic Health record and Hospital information Systems

The proper design of the repository of clinical data and its integration with the EHR has a twofold requirement. On the one hand, resorting to basic principles for the organization of the data to be collected for the Fit4MedRob project, on the other hand the choice of interoperability standards for the integration of the repository with the local EHR. The proposal of this deliverable is to use as a reference for the design of the data repository the FAIR guiding principles, and use, when possible, the OMOP/OHDSI framework for the implementation of the repository, and to apply HL7 FHIR as a tool to establish the communication between the repository and the local EHR.

The FAIR Guiding Principles[26], which specify criteria for Findability, Accessibility, Interoperability, and Reusability of data, have gained increasing attention of the research community (Table 1). To both contribute to and benefit from an expanding FAIR ecosystem, research projects should be committed to delivering a data management and software infrastructure that follows the FAIR Principles, in order to increase the benefit of the data, information, and knowledge captured throughout the project.

The FAIR Principles specify criteria for data as well as metadata ("data about data"). It concerns all kinds of descriptions of data, or more broadly, of digital objects, such as authorship, access conditions, or usage statistics.

FINDABILITY	
F	F1. (Meta)data are assigned a globally unique and persistent identifier F2. Data are described with rich metadata (defined by R1 below) F3. Metadata clearly and explicitly include the identifier of the data they describe F4. (Meta)data are registered or indexed in a searchable resource
ACCESSIBILITY	
A	A1. (Meta)data are retrievable by their identifier using a standardized communications protocol A1.1 The protocol is open, free, and universally implementable

[26]M. D. Wilkinson *et al.*, "The FAIR Guiding Principles for scientific data management and stewardship," *Sci. Data*, vol. 3, no. 160018, 2016, doi: 10.1038/sdata.2016.18.

	A1.2 The protocol allows for an authentication and authorization procedure, where necessary A2. Metadata are accessible, even when the data are no longer available
INTEROPERABILITY	
I	I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation. I2. (Meta)data use vocabularies that follow FAIR principles I3. (Meta)data include qualified references to other (meta)data
REUSABILITY	
R	R1. (Meta)data are richly described with a plurality of accurate and relevant attributes R1.1. (Meta)data are released with a clear and accessible data usage license R1.2. (Meta)data are associated with detailed provenance R1.3. (Meta)data meet domain-relevant community standards

Table 1. FAIR Guiding Principles for Scientific Data Management and Stewardship

To deliver a data management and software infrastructure that adheres to the FAIR (Findability, Accessibility, Interoperability, Reusability) principles, the adoption of standard terminologies, data models, and APIs is a key step. Addressing the FAIR principles right from the start and throughout the development of a technological intervention rather than upon completion, influences design and implementation decisions that are made. This will remove the burden of data conversion at a later stage, and of software redesign.

Among the initiatives that support the application of the FAIR principles, the Observational Health Data Sciences and Informatics (OHDSI) collaboration has put in place a collaborative effort to standardize terminologies and develop a homogeneous data structure for observational healthcare data[27]. OHDSI (https://www.ohdsi.org/) is an open science initiative involving an international network of researchers and data partners, who focus on methodological research,

[27] Hripcsak G, et al. Observational health data sciences and informatics (OHDSI): opportunities for observational researchers. Stud Health Technol Inform 2015;216:574–578

open-source analytics, and clinical applications to advance the generation and dissemination of reliable medical evidence from observational data[28].

The OHDSI community has adopted the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM), an open community standard for representing the structure and content of observational data[29] [30]. OMOP CDM improves interoperability by standardizing both data structure and language, and enabling the sharing of analytics methods. The OHDSI collaborative network, common data model and library of analytics methods present an opportunity to make analytics techniques potentially available to all the data partners of the OHDSI network[31].

Observational data provides a view of what happens to a patient while receiving healthcare. Data is collected for different purposes: (i) to support the conduct of healthcare through the EHR, (ii) to manage the payment for healthcare (iii) to facilitate research for example using registry data. All three are routinely used for clinical research, the first two as secondary use data, and all three typically have their unique formatting and encoding of the content.

Research results must be drawn from many data sources. In order to do that, data needs to be harmonized into a common data standard. In addition, patient data requires a high level of protection, strict data use agreements and complex access control. A common data standard can alleviate this need by omitting the extraction step and allowing a standardized analytic to be executed on the data in its native environment - the analytics comes to the data instead of the data to the analytics, as shown in Figure 3.d.1.

This standard is provided by the Common Data Model (CDM). The CDM, combined with its standardized content, will ensure that research methods can be systematically applied to produce meaningfully comparable and reproducible results.

An overview of all the tables in the CDM is provided in Figure 7.

[28] Reps JM, et al. Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data. J Am Med Inform Assoc. 2018 Aug 1;25(8):969-975. doi: 10.1093/jamia/ocy032. PMID: 29718407; PMCID: PMC607783

[29] Overhage JM, et al. Validation of a common data model for active safety surveillance research. J Am Med Inform Assoc 2012;19(1):54–60.

[30] Voss, EA, et al. Feasibility and utility of applications of the common data model to multiple, disparate observational health databases. J. Am. Med. Inf. Assoc. 22, 553–564 (2015).

[31] Reps JM, et al. Feasibility and evaluation of a large-scale external validation approach for patient-level prediction in an international data network: validation of models predicting stroke in female patients newly diagnosed with atrial fibrillation. BMC Med Res Methodol. 2020 May 6;20(1):102. doi: 10.1186/s12874-020-00991-3. PMID: 32375693; PMCID: PMC7201646.

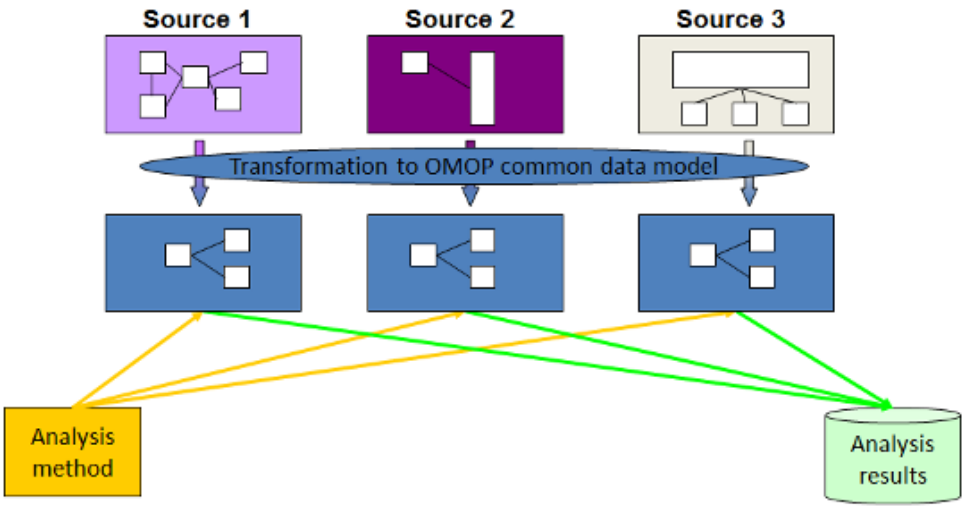


Figure 7. The overall architecture of OMOP

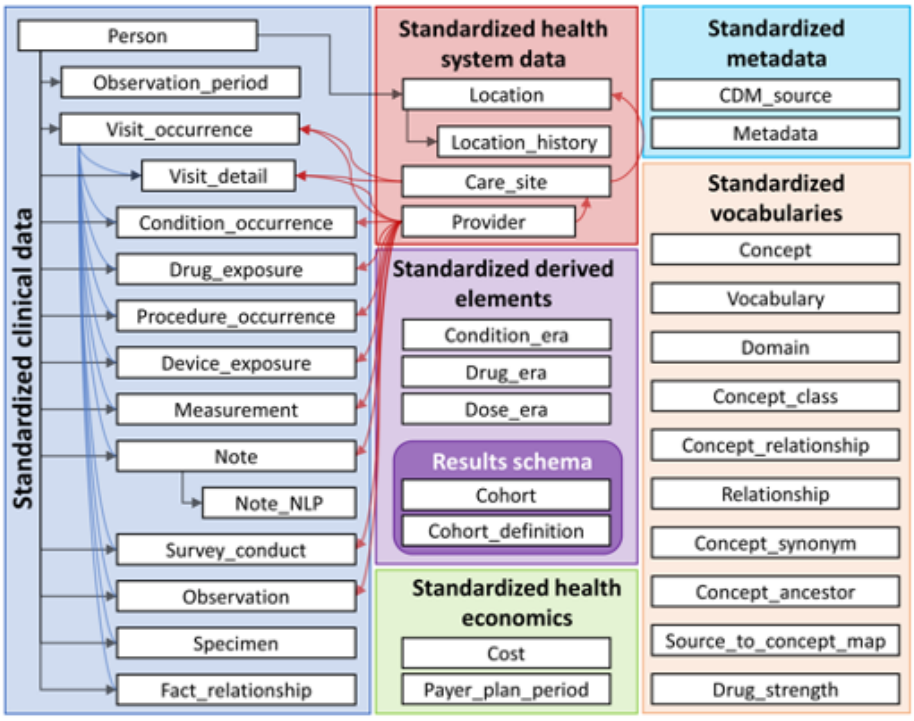


Figure 8: OMOP - The Common Data Model

The CDM is optimized for typical observational research purposes of

- Identifying patient populations with certain healthcare interventions (drug exposure, procedures, healthcare policy changes etc.) and outcomes (conditions, procedures, other drug exposures etc.),
- Characterization of these patient populations for various parameters like demographic information, disease natural history, healthcare delivery, utilization and cost, morbidities, treatments and sequence of treatment etc.,
- Predicting the occurrence of these outcomes in individual patients,
- Estimating the effect these interventions have on a population

To achieve this goal, the development of the CDM follows the following design elements:

- **Suitability for purpose**: The CDM aims to provide data organized in a way optimal for analysis, rather than for the purpose of addressing the operational needs of health care providers or payers.
- **Data protection**: All data that might jeopardize the identity and protection of patients, such as names, precise birthdays etc. are limited
- **Design of domains**: The domains are modeled in a person-centric relational data model, where for each record the identity of the person and a date is captured as a minimum. Here, a relational data model is one where the data is represented as a collection of tables linked by primary and foreign keys.
- **Rationale for domains**: Domains are identified and separately defined in an entity-relationship model if they have an analysis use case (conditions, for example) and the domain has specific attributes that are not otherwise applicable. All other data can be preserved as an observation in the observation table in an entity-attribute-value structure.
- **Standardized Vocabularies**: To standardize the content of those records, the CDM relies on the Standardized Vocabularies containing all necessary and appropriate corresponding standard healthcare concepts.
- **Reuse of existing vocabularies**: If possible, these concepts are leveraged from national or industry standardization or vocabulary definition organizations or initiatives, such as the National Library of Medicine, the Department of Veterans' Affairs, the Center of Disease Control and Prevention, etc.
- **Maintaining source codes**: Even though all codes are mapped to the Standardized Vocabularies, the model also stores the original source code to ensure no information is lost.
- **Technology neutrality**: The CDM does not require a specific technology. It can be realized in any relational database, such as Oracle, SQL Server etc., or as SAS analytical datasets.
- **Scalability**: The CDM is optimized for data processing and computational analysis to accommodate data sources that vary in size, including databases with up to hundreds of millions of persons and billions of clinical observations.
- **Backwards compatibility**: All changes from previous CDMs are clearly delineated in the github repository (https://github.com/OHDSI/CommonDataModel). Older versions of the CDM can be easily created from the current version, and no information is lost that was present previously.

Events of different nature are organized into Domains. These Events are stored in tables and fields which are Domain-specific, and represented by Standard Concepts that are also Domain-specific as defined in the Standardized Vocabularies. Each Standard Concept has a unique Domain assignment, which defines which table they are recorded in. Even though the correct Domain assignment is subject for debate in the community, this strict Domain-table-field correspondence rule assures that there is always an unambiguous location for any code or concept. There are a total of 30 Domains, as shown in the table below..

Concept Count	Domain ID	Concept Count	Domain ID
1731378	Drug	183	Route

477597	Device	180	Currency
257000	Procedure	158	Payer
163807	Condition	123	Visit
145898	Observation	51	Cost
89645	Measurement	50	Race
33759	Spec Anatomic Site	13	Plan Stop Reason
17302	Meas Value	11	Plan
1799	Specimen	6	Episode
1215	Provider Specialty	6	Sponsor
1046	Unit	5	Meas Value Operator
944	Metadata	3	Spec Disease Status
538	Revenue Code	2	Gender
336	Type Concept	2	Ethnicity
194	Relationship	1	Observation Type

Table 2. List of domains and number of concepts associated with each of them.

All clinical events in the OMOP CDM are expressed as concepts, which represent the semantic notion of each event. They are the fundamental building blocks of the data records and they are stored in the CONCEPT table. An example of record is shown in Figure 9 (source https://ohdsi.github.io/TheBookOfOhdsi/StandardizedVocabularies.html)

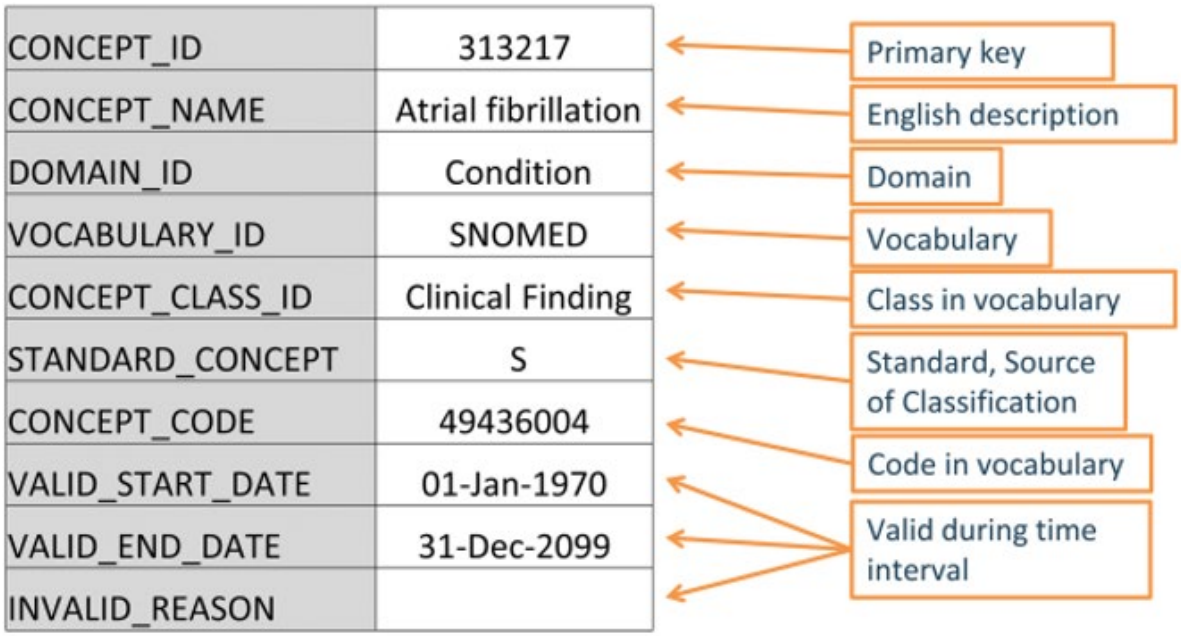


Figure 9. Concept table

Each concept is represented by a code (CONCEPT_CODE) derived by a standard vocabulary. The OMOP Standardized Vocabularies are a foundational part of the OHDSI research network, and an integral part of the CDM. They allow standardization of methods, definitions and results by defining the content of the data. Usually, finding and interpreting the content of observational healthcare data, whether it is structured data using coding schemes or laid down in free text, is passed all the way through to the researcher, who is faced with a myriad of different ways to describe clinical events. OHDSI requires harmonization not only to a standardized format, but also to a rigorous standard content.
All vocabularies of the Standardized Vocabularies are consolidated into the same common format. This relieves the researchers from having to understand and handle multiple different formats and life-cycle conventions of the originating vocabularies. All vocabularies are regularly updated. In order to obtain the Standardized Vocabularies it is possible to use a resource called ATHENA (http://athena.ohdsi.org).
Each vocabulary has a short case-sensitive unique alphanumeric ID, which generally follows the abbreviated name of the vocabulary, omitting dashes. For example, ICD-9-CM has the vocabulary ID “ICD9CM”. There are 111 vocabularies currently supported by OHDSI, of which 78 are adopted from external sources, while the rest are OMOP-internal vocabularies. As regards LOINC concept, they can be used in such domains as “Observation”, “Measurement”, “Procedure” and “Meas Value”.

Any two concepts can have a defined relationship, regardless of whether the two concepts belong to the same domain or vocabulary. These relationships provide translations from non-standard to Standard concepts, supported by two relationship ID pairs. The purpose of these mapping relationships is to allow a crosswalk between equivalent concepts to harmonize how clinical events are represented in the OMOP CDM.

The components of a modular distributed system must be technically able to exchange information and know how to decode it in order to use its contents appropriately, thus guaranteeing semantic interoperability. One of the ways to

easily connect different components is to use appropriate standards to manage and harmonize the data exchanged. Various standards have been proposed in the literature for specifying models in the clinical setting. Among these, the HL7 FHIR standard leverages existing logical and theoretical models to provide a simple, consistent, easy to implement and rigorous mechanism for exchanging data between healthcare applications in the form of resources. All FHIR resources share a common way of being technically defined and represented. They are built starting from data types that define models of generic and reusable elements. The resources are modeled on a wide range of concepts related to health, both clinical (such as demographics, health conditions and treatments related to patients) and administrative (such as professionals, organizations and places).

The HL7 FHIR R4 specification is organized into five levels shown in Figure 3.d.4, each of which represents a different functional area. The most relevant ones in the context of the project are levels 3 and 4. Level 3 includes a link to real-world concepts in the healthcare system explained through the *Administration* module, which deals with the base data of the people (patients and doctors), services and organizations involved. Level 4 is aimed at record-keeping and exchanging data for the healthcare process. The salient modules of level 4 are the *Clinical* category (relating to the fundamental information of a patient documented by healthcare providers during the course of clinical care), *Diagnostics* (which deals with the reporting of clinical diagnostics, including laboratory tests and symptoms) and *Medications* (for prescribing and administering drugs).

FHIR supports three standard formats for data exchange: XML (eXtensible Markup Language), JSON (JavaScript Object Notation) and RDF (Resource Description Framework). These types are among the most widespread in the computer science field and can be interpreted by any system, regardless of the tools used to develop it. FHIR therefore meets the needs of both end users and developers, allowing them to search and access information quickly. The JSON format appears to be the most used for its characteristics that combine compactness and ease of processing at a computational level, together with a discrete simplicity of reading by the human expert.

Figure 10. The five levels of the organization of HL7 FHIR R4 specification.

The decision to use the FHIR standard can facilitate the implementation of the FAIR principles in the management of the exchanged data. Adherence to the FAIR principles allows to make data accessible and reusable by third parties. The FHIR standard, in addition to specifying the content of the data exchanged between healthcare applications, defines the possible methods to manage and implement the exchange. In particular, FHIR is described as a RESTful specification. The RESTful API is a generic interface that can be used to push and pull data between systems. This is a client/server API designed to follow RESTful design principles for CRUD operations, along with search support, using the HTTP request/response model. Client applications can then use the FHIR resources to represent the patient's clinical information and send it to the server, communicating via REST.

APIs or Software Development Kits (SDKs) for HL7 FHIR have been made available, which facilitate implementing specification-compliant services. Examples of such APIs are:

- HAPI FHIR [31], a free and open source Java API
- SMART on FHIR [32], which provides libraries for a variety of programming lan-guages, including Python and Swift (IOS).

FIRELY .NET SDK [33], the official support SDK for working with HL7 FHIR on the Microsoft .NET (dotnet) platform.

5.5 The Health care-provider clinical dashboard

The dashboards should allow clinicians to quickly visualize actionable data to inform and optimize clinical and organizational performance, to monitor the progress of the rehabilitation activities carried out by the patient by selectively and at different levels of detail, displaying a set of performance indicator parameters.
Dashboards are typically embedded in complex healthcare organizations with massive data streams and end users with distinct needs. Thus, designing effective dashboards is a challenging task: even the concept of the dashboard remains ill-defined. The design of an effective dashboard concerns researchers, informaticists, clinical managers, and healthcare administrators. Evaluation of the dashboard should take into consideration usability and cognitive load issues, together with effectiveness.
A recent review[32] analyzes the vast published literature on healthcare “dashboards”. Requirements and challenges of hospital dashboards are analyzed in another review[33].

Functional requirements for both quality and clinical dashboards include the following:
Customization: It enables users to change the type of indicators displayed by the dashboard and optimize their view of the content shown on the screen to best suit their needs and preferences.
Alerts creating: This feature creates alerts for indicators that demand real-time monitoring. It also gives an alert when the value of the indicators exceeds the defined standard.
Tracking: This feature might track the location of patients for real-time monitoring of ongoing activities.
Measuring performance indicators: This involves comparing indicators with standards or the national average and comparing indicators over time.
Reporting: The ability to prepare visual reports based on clinical performance indicators in clinical dashboards (see next, “Data analytics”). It also involves the ability to create output files in various formats (Excel, Word, PDF).

[32] Helminski D, Kurlander JE, Renji AD, Sussman JB, Pfeiffer PN, Conte ML, Gadabu OJ, Kokaly AN, Goldberg R, Ranusch A, Damschroder LJ, Landis-Lewis Z. Dashboards in Health Care Settings: Protocol for a Scoping Review. JMIR Res Protoc. 2022 Mar 2;11(3):e34894. doi: 10.2196/34894. PMID: 35234650; PMCID: PMC8928055.

[33] Rabiei, R., Almasi, S. Requirements and challenges of hospital dashboards: a systematic literature review. BMC Med Inform Decis Mak 22, 287 (2022). https://doi.org/10.1186/s12911-022-02037-8

5.5.1 Data Analytics

Data Analytics is the science of analysis—using data for decision-making. Analytics involves the use of data, analysis, and modeling to arrive at a solution to a problem or to identify new opportunities.

Figure 11: Data is ultimately used to make a decision and take a course of action, from [34]

All the techniques that compute metrics and indicators from the data acquired during the delivery of the interventions (see Section 2 for a detailed analysis of all the possible data) are within this scope.
Metrics are quantitative measurements e.g., blood pressure is a metric that can be used by an individual to measure some aspects of cardiovascular performance/quality. However, metrics alone are not sufficient; we need to tie a metric to a target goal to determine whether a certain desirable goal has been attained. Metrics tied to a certain target are called indicators; indicators are markers for progress or achievement. Indicators can be consolidated on a screen using visualization tools such as figures, charts, colors, or numbers. These indicators are displayed in a simple-to-use and easy-to-understand way on dashboards.
Such visualizations are typically bar plots, area plots, and tables, using colors to highlight important information. Given the huge amount of information nowadays collected and available, data visualization techniques are developed to explore and navigate data and results in more efficient and natural ways. A detailed discussion about data visualization techniques is out of the scope of this document.

5.6 The patient dashboard, coaching and message handling

The patient dashboard is a key feature of the platform, designed to empower patients to take an active role in managing their own health. Through the dashboard, patients can access a clear and intuitive interface that allows them to visualize their data over time. This data can include information on the movements, heart rate, respiratory rate, sleep analysis, etc.. By reviewing this data, patients can gain insights into their overall health and identify areas where they may need to make lifestyle modifications.
The patient dashboard also includes a system of alerts and recommendations that can help patients stay on track with their health and rehabilitation goals. For example, if the dashboard detects that a patient's heart rate has been consistently elevated over a period of time, it may recommend that the patient engage in stress-reducing activities or seek medical attention. Similarly, if the dashboard identifies that the patient's dietary habits are suboptimal, it may provide suggestions for dietary changes that could improve the patient's overall health.

The dashboard also enables patients to communicate with their doctors and caregivers directly. If the patient experiences any pain or discomfort during their rehabilitation, they can use the dashboard to send a message to their

[34] El Morr, C., Ali-Hassan, H. (2019). Healthcare, Data Analytics, and Business Intelligence. In: Analytics in Healthcare. SpringerBriefs in Health Care Management and Economics. Springer, Cham. https://doi.org/10.1007/978-3-030-04506-7_1

doctor for advice on how to manage the pain. The doctor can also use the dashboard to monitor the patient's progress and adjust their treatment plan as needed.

6 Prescription of the robotic intervention

6.1 Prescription system

6.1.1 Decision support systems, Guidelines and Decision aids

Clinical practice guidelines are widely applied tools to improve the standard quality of care, patient access, treatment outcomes, appropriateness of care and achieve cost containment thanks to recommendations compiled by experts on the basis of the available evidence. The development of computerized clinical guidelines has the goal to make guidelines readily available at the point of care. Their implementation within an EHR promises to achieve an "evidence-based" decision support system able to provide suggestions while collecting data and to assess the patients' status dynamically. Given the limited evidence available about the effectiveness of robotic interventions in rehabilitation, there are currently no guidelines that can be readily used. The analysis of the Italian guidelines for Medical Physics and Rehabilitation, however, (Santilli V., 2017), suggests that there is room to associate a robotic intervention to the assessment of specific status of the patients.

DISEASE	Recommendations	References
Stroke, Parkinson, Multiple Sclerosis Parkinson's disease	On the basis of the patients' status (evaluated using a quantitative scale) defined personalized rehabilitation programs.	*-Smania N. et al., 2009.* *-Hemphill J.C. et al., 2015;* *-Barry G., Galna B., Rochester L., 2014.* *-Abbruzzese G. et al., 2016.* *-Singleton J.R. et al., 2015.*
Stroke	Perform aerobic interventions.	*-Hemphill J.C. et al., 2015;*

Multiple Sclerosis	recreation and relaxation interventions are advisable	*-Fragoso Y.D. et al., 2014 Aug;* *-Fragoso Y.D. et al., 2014 Aug* *-Smania N. et al., 2009.* *-Bennett S. et al., A Practical Guide to Rehabilitation in Multiple Sclerosis.*
Parkinson's disease	Virtual reality and exergames are effective interventions	*-Ferreira J. et al., 2013.* *-Keus S.H. et al., 2007.*

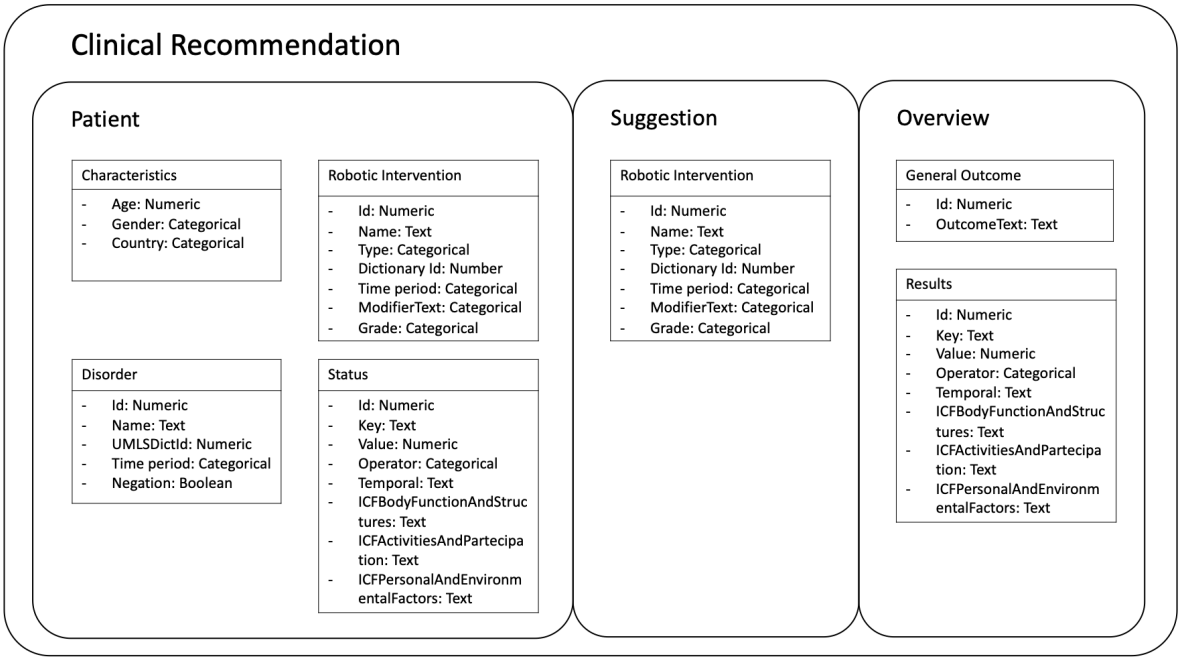
From the ICT architecture of robotic interventions, it seems advisable to propose that the prescription is supported by a so-called "decision-aid" rather than a fully formalized decision support system. In particular, the guiding principle can be to assess the status of the patient resorting to ICF and the data model defined in the previous section, and then to manually select the type of robotic intervention as well as its settings within a specified range of options dependent on the patients' status. The patient status will incorporate also a psychophysical assessment to integrate the information provided by the aforementioned standards. This will be achieved using dedicated AI-based tools to detect patients' stress levels which are a common problem among patients undergoing rehabilitation and can be caused by a variety of factors. By detecting stress in real-time, we can adjust the rehabilitation program to make it more effective and reduce frustration[35] .

Moreover, stress is a significant risk factor for a wide range of health issues, and by providing personalized stress management techniques and recommendations, we can help patients improve their overall health and well-being[36]. For instance, if we notice that a patient is experiencing high levels of emotional stress, we can modify the therapy to include relaxation techniques such as meditation or breathing exercises.

The overall conceptual scheme is represented by the figure reported below. Patients' data define the state of the subject, the suggestion collects the data related to the robotic intervention and the outcome reports the results of the interventions. This type of simple, high level representation of the decision process is amenable to perform analytics and, in the long run, to define a learning system based on the collected cases and data.

[35] Bonarini, Andrea, et al. "Stress recognition in a robotic rehabilitation task." Robotic helpers: user interaction, interfaces and companions in assistive and therapy robotics, a workshop at ACM/IEEE HRI. 2008.

[36] Toussaint, Loren, et al. "Effectiveness of progressive muscle relaxation, deep breathing, and guided imagery in promoting psychological and physiological states of relaxation." Evidence-Based Complementary and Alternative Medicine 2021 (2021).



6.1.2 Modules to set PRI (Progetto Riabilitativo Individuale) and the pri (programma riabilitativo individuale)

An essential aspect of the translation of robotic interventions into practice is to be able to consider them as components of the standard rehabilitation strategies.

The definition of a rehabilitation strategy should include the following aspects: taking charge of the patient, patients' evaluation, the elaboration of a rehabilitation project and the implementation of a specific intervention program.

Among the organizational innovations, the DGR 3111/2006 has provided, also for the services provided in an outpatient specialist regime, the definition of a personalized plan of rehabilitation for each user through the use of two tools:

- The Individual Rehabilitation Project – PRI
- The Individual Rehabilitation Program – pri

The Individual Rehabilitation Project (PRI) represents the specific, synthetic and comprehensive tool that covers all these aspects, unique for each person, defined by the rehabilitation specialist and shared with the other professionals involved. The PRI, applying the parameters of impairment, activity and participation company listed in the ICF coding system, defines the prognosis, expectations and priorities of the patient and his/her family members; it is shared with the patient and, when possible, with the family and caregivers; it defines the characteristics of congruity and appropriateness of the different interventions, as well as the conclusion of the taking into health care in relation to the outcomes achieved.

The pri defines the procedures for taking care of the patient, the areas of intervention and the immediate and/or short-term goals, updating them over time. It defines methods and times for the delivery of individual services, outcome measures appropriate for the evaluation of interventions, the expected outcome based on these measures and the verification time of the achievement of a given outcome; it identifies individual operators involved in the interventions and defines the related commitment; it constitutes an element verification of the PRI. Robotic interventions seem therefore to be suitable to represent a crucial component of specific sets of pri.

The purposes, contents, characteristics and methods of use of the two tools are specified in detail in the Guidelines of the Ministry of Health for the activities of rehabilitation, OJ 30 May 1998, general series - n.124 - Provision 7 May 1998.

6.2 Integration with the Electronic Health record and Hospital information Systems

The integration of the overall architecture with the hospital information system and the Electronic Health record is a complex task that may need to face technical and regulatory challenges. In principle there are three different strategies that can be pursued. In the first strategy, data are imported into the Fit4MedRob system from the EHR for evaluation of the patient's case and for performing the prescription. In the second scenario, data is collected by the Fit4MedRob system and then imported into the EHR. Finally, in the third option, which is the most desirable, communication is bidirectional. Although the three cases are possible, in the deliverable we will analyze only the first scenario, with the idea that Fit4MedRob will work independently from the different EHR configurations and expectations.
Under the assumption that the Fit4MedRob data repository supporting decision making is based on a relational model, such as OMOP (see previous sections) the integration with the local EHR can be thus performed by collecting the needed information by querying the EHR with an interoperability communication standard, such as FHIR.

7 Security and standards

This chapter addresses two central issues for the definition of the architecture described in the previous Chapter 1: (i) **security** in the management and transmission of clinical data (as well as their privacy, but on this topic we also refer to the work of Mission 1 of Fit4MedRob), (ii) and the most common communication standards for achieving effective **interoperability** of clinical data across different structures.

The security standards and protocols for medical rehabilitation in general, and therefore for robotic rehabilitation as well, focus on protecting sensitive data of the individuals involved in trials and diagnostic and rehabilitation pathways, such as medical conditions, medical history, and other personal details. This means that communications between healthcare professionals, patients, and other healthcare operators must be protected by security measures and protocols, such as cryptographic, authentication, and authorization protocols. Data protection is in fact a critical aspect of medical rehabilitation and tele-rehabilitation, as it involves the processing of sensitive personal and health data. Here are some key technical considerations for data protection in this context:

1. Secure communications: All communication channels, including email and instant messaging, should be secured to protect them from interception and unauthorized access. This can be achieved using encryption and secure messaging protocols. In more details, all data transmitted between the rehabilitation devices and the remote server should be encrypted to protect it from interception and unauthorized access. Shortly, encryption is the process of converting plain text data into a coded form, known as ciphertext, to prevent unauthorized access to the data. In the medical context, encryption is an essential security measure to protect sensitive patient data during transmission over networks, such as the internet. There are several encryption algorithms that can be used to protect data in transit, such as:
    a. Advanced Encryption Standard (AES): AES is a widely-used encryption algorithm that is considered to be very secure. It uses a symmetric key encryption method, which means that the same key is used to encrypt and decrypt the data.
    b. Triple Data Encryption Standard (3DES): 3DES is an encryption algorithm that uses three rounds of encryption to provide a high level of security. Like AES, it uses a symmetric key encryption method.
    c. Secure Sockets Layer (SSL)/Transport Layer Security (TLS): SSL and TLS are protocols that are used to secure communication between devices over the internet. They use a combination of symmetric and asymmetric key encryption methods to provide a high level of security.

    Encryption can be implemented at various levels in a tele-rehabilitation system, including:
    a. Application layer encryption: Encryption can be implemented within the tele-rehabilitation application itself to encrypt data before it is transmitted over the network.

b. Transport layer encryption: Encryption can be implemented at the transport layer using protocols such as SSL/TLS to encrypt data during transmission.
c. Network layer encryption: Encryption can be implemented at the network layer using virtual private networks (VPNs) to provide a secure tunnel for data transmission.

Also, secure file transfer protocols, such as SFTP (Secure File Transfer Protocol) and FTPS (FTP over SSL/TLS), should be always used to encrypt data in transit to protect it from interception or tampering. In addition to choosing an appropriate encryption algorithm and level, it is important to manage encryption keys properly. Encryption keys are used to encrypt and decrypt the data, and they must be kept secure to prevent unauthorized access. Best practices for key management include using strong, complex passwords and regularly changing keys.

2. Access controls, a key component of information security that help to ensure that only authorized users are able to access sensitive data or resources. Access controls should be implemented to restrict access to sensitive data to authorized personnel only. They typically involve a combination of technical, physical, and administrative measures. Some common examples of access controls include:
   a. Passwords: Passwords are one of the most basic and widely used forms of access control. Users are typically required to enter a username and password to access a system or resource.
   b. Two-factor authentication (2FA): 2FA is a stronger form of authentication that requires users to provide two forms of identification, such as a password and a fingerprint or a security token.
   c. Role-based access control (RBAC): RBAC is a method of restricting access to resources based on the user's role within an organization. For example, a user might have access to only the resources that are necessary to perform their job duties.
   d. Access control lists (ACLs): ACLs are used to define which users or groups have access to specific resources. They can be used to grant or revoke permissions to resources as needed.
   e. Physical access controls: Physical access controls, such as security cameras, keycard readers, and biometric scanners, are used to restrict access to physical locations, such as data centers or server rooms.
   f. Administrative controls: Administrative controls, such as policies, procedures, and training, are used to ensure that access controls are implemented and enforced effectively.

   An effective implementation of access controls, clearly involves conducting a risk assessment to identify potential threats and vulnerabilities, hence implementing appropriate technical and administrative measures to address these risks.
3. Data minimization: Only the minimum amount of data necessary for the rehabilitation process should be collected and processed. This can help to reduce the risk of data breaches and ensure compliance with data protection regulations. Also, organizations should have policies and procedures in place for deleting personal data when it is no longer necessary for the purpose for which it was collected. Finally, Regular data audits can help organizations to identify and remove unnecessary or outdated personal data, and ensure that they are complying with data minimization principles.
4. Data anonymization and pseudonymization, that are techniques used to protect the privacy of individuals by removing or masking their identifying information.

   Data anonymization involves the removal of all personally identifiable information from a dataset, so that the data can no longer be linked back to a specific individual. This can be achieved by removing names, addresses, birth dates, social security numbers, and any other information that could be used to identify an individual. Once the data has been anonymized, it can be used for research, analysis, or other purposes without compromising the privacy of the individuals in the dataset.

   Pseudonymization, on the other hand, involves replacing identifying information with a pseudonym, or a random identifier. This allows the data to be used for analysis while still protecting the privacy of the individuals in the dataset. For example, a patient's name and address might be replaced with a randomly generated ID number. Pseudonymization can be useful for research or analysis that requires identifying trends or patterns in the data, but does not require access to the individual's personal information.

It's worth noting that while both data anonymization and pseudonymization can be effective in protecting privacy, neither technique is foolproof. For example, anonymized data can sometimes be re-identified by combining it with other datasets, while pseudonymized data can be linked back to individuals if the pseudonyms are not kept secure. Therefore, it's important to carefully consider the risks and benefits of each technique in the specific context in which they are being used, and to implement appropriate technical and organizational measures to ensure the security of the data.

5. Data backup and recovery: Regular backups of data should be taken to ensure that it can be recovered in the event of a data loss or corruption. These backups should be stored securely and offsite to protect them from physical damage or theft.
6. Vulnerability management: Regular vulnerability assessments and penetration testing should be conducted to identify and mitigate potential security risks. Any vulnerabilities that are identified should be addressed promptly to reduce the risk of data breaches.
7. Monitoring and logging: All access to sensitive data should be logged and monitored to detect and respond to any unauthorized access or suspicious activity. This can help to identify potential security breaches and improve the effectiveness of security measures.

By implementing these technical measures, providers can ensure that patient data is protected from unauthorized access and use, and comply with data protection regulations such as the General Data Protection Regulation (GDPR) in the EU and the Health Insurance Portability and Accountability Act (HIPAA) in the United States. Overall, it's important to note that compliance with these standards is not a one-time event, but an ongoing process that requires continuous monitoring, testing, and improvement to ensure the security of sensitive information.

Regarding communication standards, one of the most widely used in the medical field is the Health Level Seven International (HL7)[37] protocol. HL7 is a non-profit international organization (based in the USA) that standardizes the exchange of healthcare information through the use of communication standards.

HL7 has developed a series of communication standards for the exchange of healthcare information between different healthcare information systems, using a client-server architecture to exchange data reliably and securely, addressing interoperability goals. Its standards have been adopted worldwide and have been used to develop numerous healthcare information systems. The main HL7 standards include:

1. HL7 Version 2.x - This is the primary communication standard used for the exchange of healthcare information. This standard was developed to facilitate the exchange of information between different healthcare information systems.
2. HL7 Version 3 - This is a newer communication standard that uses an object-oriented methodology. This standard aims to simplify interoperability between healthcare systems and increase the flexibility and efficiency of information exchange.
3. Clinical Document Architecture (CDA) - This is an HL7 standard for the creation and exchange of electronic clinical documents. CDA defines a standard format for electronic health documents, such as laboratory reports, imaging reports, and medication administration documents.
4. Fast Healthcare Interoperability Resources (FHIR) - This is an HL7 standard based on RESTful APIs that simplifies interoperability between healthcare systems. FHIR enables access to healthcare information in a more efficient and flexible way than previous standards.

HL7 also offers a wide range of tools and resources to help healthcare professionals and developers implement its standards. These tools include the HL7 testbed, which allows healthcare professionals to test the implementation of their healthcare information systems using HL7 standards, and the conformance testing service, which helps developers verify their systems' compliance with HL7 specifications.

[37] https://www.hl7.org/

Another important communication standard for medical rehabilitation is the Integrating the Healthcare Enterprise (IHE)[38] protocol. Healthcare Enterprise (IHE) is a global non-profit initiative that promotes interoperability among healthcare information systems. Founded in 1998 by the Radiological Society of North America (RSNA) and Healthcare Information and Management Systems Society (HIMSS), IHE focuses on harmonizing protocols, standards, and information exchange formats to improve healthcare efficiency.

IHE develops interoperability profiles that describe how different healthcare information systems can communicate with each other. These profiles are based on existing standards and specify how to use them effectively to address healthcare interoperability needs. IHE's interoperability profiles include:

1. Cross Enterprise Document Sharing (XDS) - Enables secure exchange of medical documents across different healthcare organizations.
2. Patient Identifier Cross-Referencing (PIX) - Provides a standard method for uniquely associating patient records across different healthcare organizations.
3. Retrieve Form for Data Capture (RFD) - Provides a standard method for retrieving electronically collected patient data in a format compatible with paper documentation.
4. Scheduled Workflow (SWF) - Provides a standard method for communication between healthcare information systems managing exam booking and scheduling processes.
5. Mobile Access to Health Documents (MHD) - Enables secure access to medical documents on mobile devices such as smartphones and tablets.

IHE also offers a certification program, called IHE Integration Statements, which allows healthcare information system vendors to demonstrate compliance with IHE's interoperability profiles. Healthcare information system vendors can publish an integration statement attesting to compliance with a specific IHE interoperability profile. This allows healthcare professionals to select healthcare information systems that comply with IHE specifications.

Finally, it's important to note that many healthcare organizations adopt data protection regulations, such as the European General Data Protection Regulation (GDPR) or the United States' Health Insurance Portability and Accountability Act (HIPAA), which establish standards for the collection, processing, and storage of patient data. Healthcare organizations must comply with these regulations to protect sensitive patient data and ensure their privacy. For specific aspects related to privacy, we also refer to the relevant documentation produced in the context of Fit4MedRob – Mission 1.

7.1 Delivery

The software used to control the robot must be secure and protected from unauthorized access. The internet connection used for telemedicine should be secure and protected by encryption and authentication to protect data and ensure privacy.

7.2 Monitoring

The monitoring and assessment of robotic intervention (including remote ones in telemedicine) can involve several stages, such as data analysis, progress evaluation, and outcome assessment. It is important that the data collected during the intervention is protected by security measures such as encryption and authentication, to ensure patient privacy and safety. Additionally, medical personnel should be able to monitor the intervention in real-time and intervene if necessary.

[38] https://www.ihe.net/

7.3 Prescription

The prescription of remote robotic intervention requires an accurate assessment of the patient, which can be conducted through physical examinations, laboratory tests, and diagnostic evaluations. Additionally, medical personnel must consider the risks and benefits of remote robotic intervention and discuss available options with the patient. Finally, medical personnel must prescribe the remote robotic intervention accurately, avoiding prescription errors and ensuring that the patient is aware of possible complications and side effects.

8 Inter-institutional data sharing

Inter-institutional data sharing has become increasingly important in today's data-driven world, as it enables organizations to collaborate, share insights, and make better-informed decisions. The creation of artificial data from real data by data augmentation techniques and the use of sandbox environments are two key components that can facilitate this process while maintaining data privacy and security.

Data augmentation refers to the technique of expanding datasets by generating new synthetic instances from the existing data, such as by randomization, substitution, dataset reconfiguration, or deep-learning techniques based on generative adversarial networks (GANs). The new synthetic data is locally created from the original data that is mimicking; it is completely anonymous by design and therefore secure for transfer to the external (for example to a central platform).

In the case of synthetic data creation by randomization, some of the features of the original dataset will be modified randomly, for example by adding noise or performing transformations. Randomization, in this sense, can make it more difficult to identify the original data within the augmented dataset and thus prevent the inadvertent disclosure of sensitive information. In the case of substitution, on the other hand, there will be a true replacement of sensitive information with synthetic information, that is, names, addresses, and synthetically generated identification codes. In the case of dataset reconfiguration, finally, the sensitive features of the dataset will be recombined or randomly shuffled, generating new synthetic data that are no longer immediately identifiable. Separate consideration should be given to ML techniques based on GANs, which, unlike the previously mentioned cases, leverage competing neural networks to produce synthetic data indistinguishable from the original data.

In general, however, beyond the individual technical specifics, a data augmentation procedure, in addition to helping protect sensitive data contained in a dataset, can also improve the performance of machine learning models trained on data generated by it. The creation of new synthetic samples indeed allows a model to learn a greater variety of features and become more robust with respect to data variation and perturbation. In other words, the model becomes more accurate in its ability to generalize on new data not present in the original dataset. Moreover, by increasing the size of the dataset, the risk of overfitting is reduced, thus avoiding the generalization problems usually encountered when training a model on modestly sized datasets. This approach can help improve the quality and diversity of shared data, making it more valuable for collaborative research and analysis while preserving the privacy of the patients.

From an implementation perspective, data augmentation techniques can be integrated into the platform in two possible ways: they can be an integral part of the data pre-processing pipeline, implemented within a data integration layer, or they can be provided as an external service by a second infrastructure, in the form of a synthetic data platform. Regardless of the implementation chosen, the data processing flow is still divided into four main phases, briefly described below:

1. Acquisition of the target dataset: collection of original data to be used as input for the generation of synthetic data
2. Dataset analysis: evaluation of data distribution, identification of outliers or missing data, and definition of data transformation rules (generally based on user input)
3. Synthetic data generation: generation of data according to different possible methods (randomization, machine learning techniques, non-linear transformations, etc.)
4. Synthetic data validation: evaluation of the consistency between synthetic data and original data, data quality assessment of the synthetic dataset

On the other hand, sandbox environments provide secure, isolated spaces where data can be shared, accessed, and manipulated without exposing sensitive information or affecting production systems. The confinement of original data within a sandbox based on services and APIs allows the creation of a local "data wrap", which is the initial dataset but

protected by the sandbox, that is not designed to be exported to external platform, but to be made available for remote interaction with algorithms and models launched directly from the local center, again in a secure manner. The sandbox architecture should enable training algorithms on proprietary datasets, both in traditional and federated modes, without accessing them, but by displaying the model output in a way that preserves the privacy of the involved data. The sandbox architecture system can be used to protect the privacy of sensitive data through two different types of isolation: data isolation and process isolation, in an environment that remains always controlled and secure. Regarding the first type of isolation, data isolation, all sensitive data intended for machine learning model training can be stored within the sandbox, making them accessible only to authorized processes and not the main operating system. In this way, sensitive data can be protected from potential cyberattacks or unauthorized access. In this sense, access control practices will also be functional to this approach, limiting the use of sensitive data for training purposes only to authorized processes and users. Access to the data wrapper services will be particularly bound to the use of access credentials, authentication systems, and role-based authorizations. The second level of isolation concerns resources. The sandbox architecture system will be specifically isolated from the host system and will virtualize the CPU, memory, and storage so that each process within the sandbox will have exclusive access only to the resources assigned to it. In this way, it will be possible to prevent unauthorized processes from accessing system resources and any type of violation. Within the same sandbox, network communications will also be limited so that internal processes cannot access the external network or communicate with other processes outside the sandbox. Even in the case of federated learning approaches, communication with the network will always be strictly controlled, and will only concern the cloud infrastructure for exchanging training information. This technique will be used to prevent the disclosure of sensitive data through network communications, and thus to avoid any type of data breach. Finally, the sandbox can protect proprietary datasets contained within it using ad hoc encryption techniques, such as end-to-end encryption or homomorphic encryption: in particular, with end-to-end encryption, the data will be encrypted within the sandbox, and only the authorized recipient, i.e., the machine learning model to be trained, will be able to decrypt them. In the case of homomorphic encryption, the machine learning model can be trained on encrypted data without having to decrypt sensitive data, as it will not need to access the original data.

The flow implemented by the sandbox system, regardless of the particular architecture chosen, will be specifically organized in the following points:

1. Loading of the proprietary dataset within the sandbox
2. Anonymization and "masking" of the proprietary dataset when possible
3. Authorized user access to the sandbox environment to initiate training of a model on the proprietary dataset
4. Initialization of the machine learning model within the sandbox
5. Training of the ML model on anonymized (or encrypted) data within the sandbox in a "shielded" manner
6. Returning the output of the trained model on the anonymized proprietary dataset, which can be shared with external platforms for its analytics functions.

By leveraging artificial data creation by data augmentation and sandbox environments, inter-institutional data sharing can be achieved more effectively, allowing organizations to harness the power of collective intelligence while maintaining compliance with data protection regulations and preserving trust among participating entities.

9 Use cases

9.1 IRCCS Istituto Giannina Gaslini

Premature neonates have higher rates of morbidity and mortality than full-term babies. The observation of spontaneous movements in the first three months of age might be instrumental to early identify those babies at risk of developing neuro-motor-cognitive deficits. MIMAS2 (Markerless Infant Movement Analysis System) is a simple, low-cost system of video analysis of spontaneous movements of newborns in their natural environment, based on a single standard RGB camera, without markers attached to the body. MIMAS2 was developed by the University of Genoa, in cooperation with the Italian Institute of Technology. Different from most systems of motion capture, which are designed to perform a precise kinematic analysis of individual body segments, MIMAS2 aims to obtain global indicators of the spontaneous whole-body movements specific to newborns while laying on their cradle. Movement descriptors and clinical evaluations required to be stored in flexible, efficient, and scalable database platforms. To this end, we developed

XTENS-MIMAS2 Biobank, a software platform able to integrate and manage large collections of heterogeneous biomedical data.

XTENS-MIMAS2 consists of:

- A web server application running on the Node JavaScript runtime. The web server exposes a RESTful Application Programming Interface (API) for both programmatic and human user access;
- A web user interface developed using the Backbone.js Model-View-Controller (MVC) framework;
- A PostgreSQL relational database to store data records and their descriptive metadata. Semi-structured metadata are stored in JavaScript Object Notation (JSON) format, taking advantage of the binary JSON functionalities offered by PostgreSQL;
- A data grid storage element, which contains all the files associated with registered data records. The data grid of choice is the integrated Rule-Oriented Data System (iRODS) middleware.

The general XTENS-MIMAS2 Biobank architecture is shown in Figure 12.

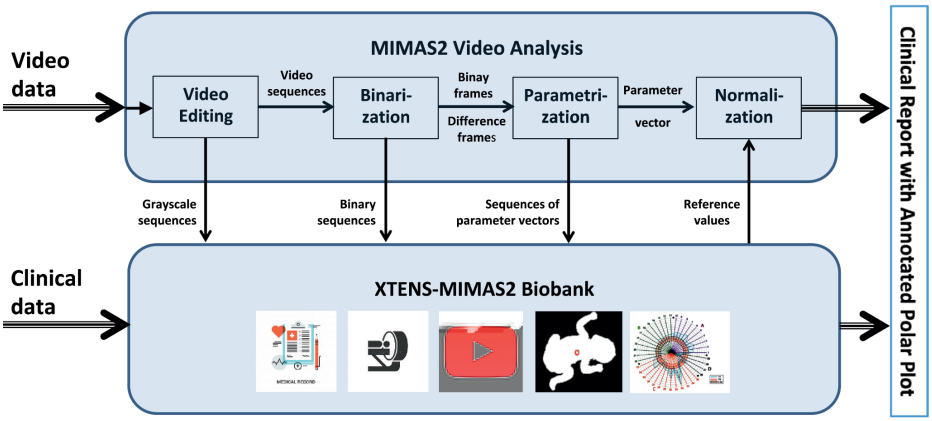


Figure 12: General MIMAS2 architecture.

Preliminary tests carried out at the Gaslini Pediatric Hospital in Genoa on 46 preterm, as cases, and 21 full-term babies, as controls, showed that XTENS-MIMAS2 Bioban was able to detect abnormal preterm babies demonstrating a potential clinical utility and a valuable computer-aided decision support system.

9.2 Eustema

In a multicenter observational study conducted between two major hospitals in Italy, involving 6,000 OCT retinal scans, we implemented an ingestion pipeline collecting the DICOM images via HL7 clients, and uploading them in a local data lake for labeling and further processing.
First, data augmentation was used to anonymize the collected OCT retina images for secure exchange between the institutions. The study involved the local collection of a large dataset of OCT scans. To ensure the privacy of the patients, the researchers applied data augmentation techniques, generating multiple variations of the original DICOM dataset. Furthermore, the data augmentation used explainability techniques in an innovative way to identify the data features that were the hardest for the model to evaluate, in order to generate data able to increase the training efficiency and the model robustness. This approach helped to protect the privacy of the patients by making it extremely difficult to identify individuals from their OCT scans. Additionally, the use of data augmentation increased the diversity of the

dataset, making it more representative and robust, thus improving the accuracy and reliability of the study's findings. By applying various data augmentation techniques such as rotation, scaling, flipping, re-combination, as well as generative adversarial networks, the researchers were able to generate additional training data, increasing the sample size and improving the robustness of the models. This resulted in more accurate and reliable segmentation of the retinal layers and clinical signs, which is critical for the diagnosis and treatment of various retinal diseases. The use of data augmentation allowed the researchers to optimize the DL models and improve their performance, making them a valuable tool for clinical practice.
The augmented dataset was exchanged between the hospitals through the use of HL7 messaging.
The application of data augmentation in this study highlights its potential as a powerful tool for anonymizing and improving the quality of large medical datasets used in research, especially in the context of DICOM and HL7 data exchange.

9.3 Campus Biomedico

The Research Unit of Computer Systems and Bioinformatics is developing a platform for telemonitoring the patient's psychophysical state during the rehabilitation process. The platform combines smart devices technology with dashboards creating a direct communication between the patients and their doctors, allowing for real-time monitoring of the patient's general health status and psychophysical stress levels. Indeed, psychological stress is a common problem among patients who are undergoing rehabilitation, and it can be caused by a variety of factors. Monitoring psychological stress is important during remote rehabilitation process as it can help healthcare providers to adapt the exercises to the patient needs to reduce frustration and improve interest[39].
The smart devices are the cornerstone of the platform, capturing a wealth of physiological data that is transmitted in real-time to the dashboard. This data includes for example heart rate, blood pressure, respiratory rate, electrodermal activity, oxygen saturation, etc. The acquisition won't be limited to physiological signals, but will also include motion-related signals from IMU sensors and information directly inserted by the patient into his/her dedicated dashboard, related to his/her lifestyle and psychological state, such as food diaries and responses to questionnaires. This feature provides doctors with additional insight into the patient's health status and allows for personalized treatment plans.
The devices communicate with the dashboard through the Bluetooth Low Energy (BLE) protocol and the data acquired by the application flows in a centralized data lake that includes also information extracted from the Health Information System of the Hospital (the Campus Bio-Medico University Hospital Foundation in our case), such as diagnostic images, EHR, clinical notes, etc. This structured and unstructured information is then cleaned, preprocessed and stored in a data repository that allows AI-based data analytics to analyze the data and detect various types of psychophysical stress. Indeed, using deep learning algorithms we can distinguish between emotional stress, physical stress, and cognitive stress by analyzing the physiological and motion data of the patient[40]. By doing so, we can provide personalized stress management techniques and recommendations to the patient.
This not only provides a decision-support and recommendation system that optimizes rehabilitation therapy but also improves the patient's lifestyle and reduces psychophysical stress. Processed data and AI results can be accessed and visualized by both the clinician and the patient. On the one hand, this allows the clinician to be aware of the patient's status and of any critical situation, to optimize therapy and intervene in a timely manner when needed by the patient. On the other hand also the patient is able to monitor and visualize this data, giving him/her a better understanding of his/her overall health and providing him/her with insights into how his/her lifestyle choices and daily habits impact his/her health. The platform also provides patients with a direct channel of communication with their doctor, allowing them to ask questions, report symptoms, and receive personalized recommendations.

The Research Unit of Advanced Robotics and Human-Centred Technologies developed an infrastructure for multimodal data management (to date implemented and tested with kinematic, temperature, heart rate and respiratory sensors) for user state estimation during human-robot interaction. The modular software architecture allows easily connecting

[39] Bonarini, Andrea, et al. "Stress recognition in a robotic rehabilitation task." Robotic helpers: user interaction, interfaces and companions in assistive and therapy robotics, a workshop at ACM/IEEE HRI. 2008.
[40] Palumbo, A.; Vizza, P.; Calabrese, B.; Ielpo, N. Biopotential Signal Monitoring Systems in Rehabilitation: A Review. *Sensors* **2021**, *21*, 7172.

different sensors or a part of them, in order to make the system adaptable to different applications. The architecture is completely ROS-based and a messaging system has been developed to connect the various devices by creating ad hoc nodes. Each component of the multimodal interface is associated with a node, so connecting or disconnecting a component can be done simply by connecting and disconnecting each device. Each node sends data on ROS topic, so that both saving and real-time processing for status estimation are allowed. The main processing unit is a node that represents the core of the architecture. The software developed in C ++ exchanges messages with the other subsystems of the platform using native nodes, which use a full-duplex channel.

The infrastructure can be adopted in different application scenarios, and it has been already used in robot-aided rehabilitation also in conjunction with virtual reality environment: robots allow to physically drive patients, multimodal sensors platform allows to estimate user state, and virtual reality can be exploited to engage patients in rehabilitation activities providing them with visual feedback to correct their actions and closing the control cycle.

The architecture has been already used in rehabilitative sessions of upper limb both in adult and pediatric patients[41] [42] [43]. In particular, in the adult application, patients were rehabilitated to a specific working activity (i.e. a lifting task at different height selected according to ISO11228-3) supported by a robotic arm, while in the pediatric application, a commercial exoskeleton for hand rehabilitation integrated with force sensors and a specifically developed serious game have been used to rehabilitate grasp force in children.

9.4 Tecnobody

In a MoveDifferent centre equipped with a full Tecnobody ecosystem (Prokin, WalkerView, D-Wall, Homing Studio, Homing Training), 20 Parkinson's patients were recruited to validate a continuum of care based on Tecnobody systems. Biomechanical assessment of the patient and motion tracking are provided by RGBD camera and load cells, thus leaving patients completely free to move without impediment. Also, HR data are recorded and displayed to the user. Firstly, patients were registered with a custom anamnestic form, stored inside the local hub. Once the profile is completed, a full evaluation is performed (Health Test, Comparative Stabilometric Test, LOS Test, Gait Analysis) to assess the global status of the patient in terms of balance, strength, locomotion, and agility. Using these objective results, the therapist is able to create a custom rehabilitation program (pri) composed by exergames, non-immersive VR and AR-VST exercises, improving the engagement. Biofeedback, quantitative and qualitative performance indices (repetitions, time of execution, reaction time, goals achieved) allow therapists to objectively monitor the patient.
All the data (test results and training assessments) are stored in the local hub, uploaded to private Cloud (accessed via token authentication) and transmitted on a HTTPS connection with a custom protocol. Uploaded data are pseudo-anonymized and GDPR compliant.

In this protocol, the first phase is performed 5 days/week inside the clinical rehabilitation centre with therapists' assistance in order to confirm the consistency between the custom rehabilitation program and patient's skills, both motor and cognitive.
Afterwards, the patient is trained by the therapist to exercise autonomously using Homing Studio (in the transfer room) with the aim of being prepared for the subsequent tele-assisted rehabilitation phase.
When a patient can continue his/her rehabilitation program at home, the therapist provides him/her a Homing system,

[41] Tamantini, C., Cordella, F., Lauretti, C., di Luzio, F. S., Bravi, M., Bressi, F., ... & Zollo, L. (2022, May). Patient-tailored Adaptive Control for Robot-aided Orthopaedic Rehabilitation. In 2022 International Conference on Robotics and Automation (ICRA) (pp. 5434-5440). IEEE.

[42] Scotto di Luzio, F., Lauretti, C., Cordella, F., Draicchio, F., & Zollo, L. (2020). Visual vs vibrotactile feedback for posture assessment during upper-limb robot-aided rehabilitation. Applied ergonomics, 82, 102950.

[43] F. Bressi, L. Cricenti, M. Bravi, F. Pannunzio, F. Cordella, M. Lapresa, S. Miccinilli, F. Santacaterina, L. Zollo, S. Sterzi, B. Campagnola, "Treatment of the Paretic Hand with a Robotic Glove Combined with Physiotherapy in a Patient Suffering from Traumatic Tetraparesis: A Case Report", Sensors 23 (7), 3484, 2023

a telerehabilitation device connected to the private cloud.
Using this connection, the patient receives rehabilitation programs remotely and at the same time the device transmits training results at the rehabilitation center (asynchronous monitoring).
The therapist can also provide real-time assistance (with a third-party software) to the patient, guiding him/her with VVoIP call (synchronous monitoring).
By means of the Tecnobody App, the patient is aware of the assigned rehabilitation activities planned and the progress recorded.
Thanks to this ecosystem, a continuum of care is realized, leaving the patient free to rehabilitate at home, maintaining a single treatment per week at the clinic.

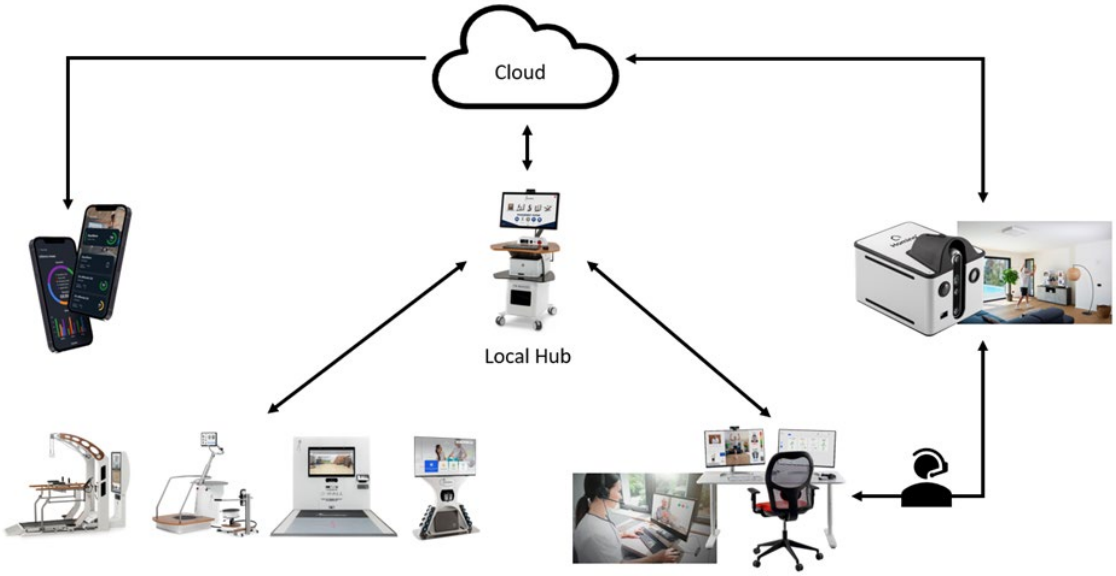


Figure 13 Data flow management diagram

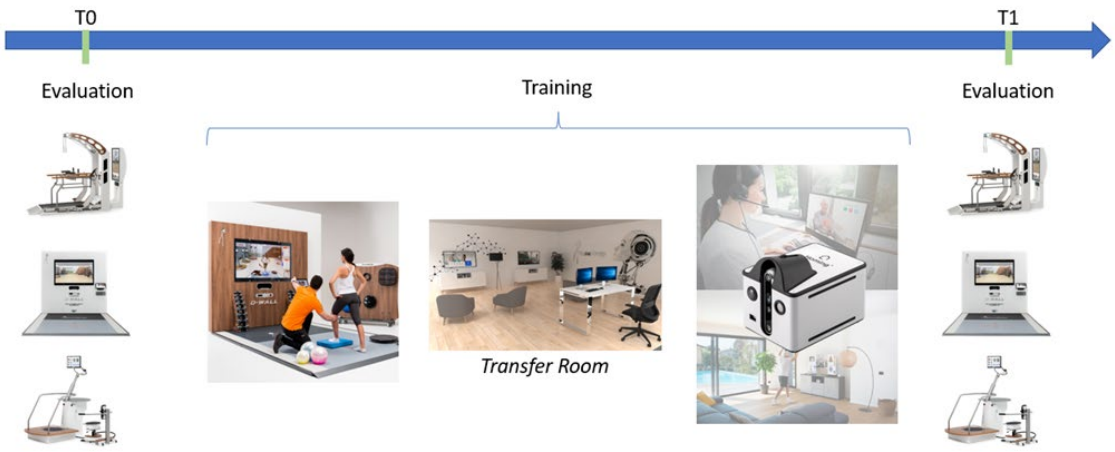


Figure 14 Use case scenario: T0 is the evaluation after patient discharge. T1 is evaluation after training program

9.5 ICS Maugeri

In ICS Maugeri is ongoing the "Palestra Digitale" ("Comprehensive digital rehabilitation") project with the aim of managing paperless the patient's rehabilitative journey[44].
At the present time the different software involved, specialized for each specific function in the clinical care pathway management, are partially integrated in the Electronic Health Record - EHR (Cartella Galileo and 4C, Dedalus Italia spa). All the software is integrated with the ADT and reports are sent to the EHR, but there is a minimal integration among the EHR and the robotic intervention.

In Figure 15 below are presented synthesized the rehabilitation phases and the main software systems involved. The patient's journey involves a circular and retroactive process; for simplicity reasons here is represented as linear.

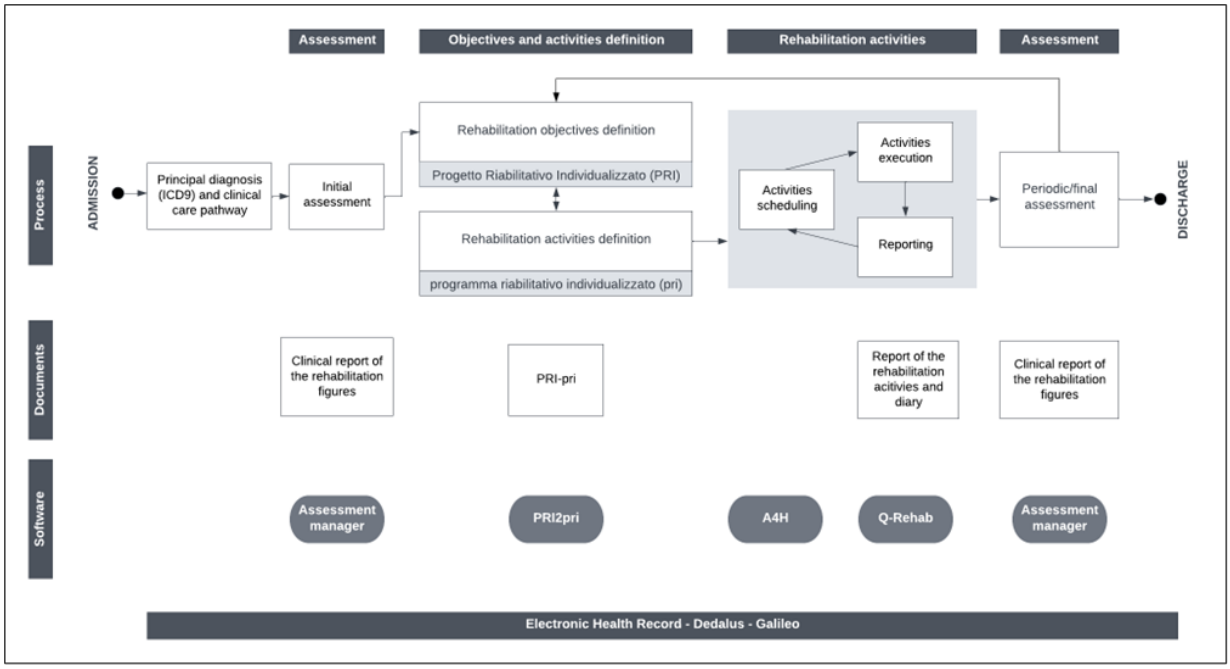


Figure 15 Rehabilitation phases and the main software systems involved.

[44] Giardini A, Traversoni S, Garbelli C, Lodigiani A. [Digitalisation and clinical care pathways in rehabilitation medicine: possible integration from the goal-planning and the rehabilitation programme design to the evaluation of clinical outcomes] [Original article in Italian]. G Ital Med Lav Erg. 2018; 40: 22-29.

De Nardi P, Giorgi G, La Manna A, Traversoni S, Giardini A. [Towards a baseline ontology for information technology in rehabilitation medicine: the nomenclator of the rehabilitation procedures]. [Original article in Italian]. G Ital Med Lav Erg. 2019; 41: 112-116.

Cardellini M, De Nardi P, Dodaro C, Galatà G, Giardini A, Maratea M, Porro I. Solving Rehabilitation Scheduling problems via a Two-Phase ASP approach. TPLP (2023 in press).

Rehabilitation objectives and activities definition

At the beginning of the patient's treatment, the physician, together with the rehabilitation team, defines the rehabilitation objects (PRI – Progetto Riabilitativo Individuale) and the rehabilitation activities (pri – programma riabilitativo individuale) for the different rehabilitation figures (software PRI2pri – BCS s.r.l. Biomedical Computering Systems).

The rehabilitation objectives are linked to the WHO-ICF classification system through a shared ontology.

The following two figures represent a real life example of the software PRI2pri adoption.

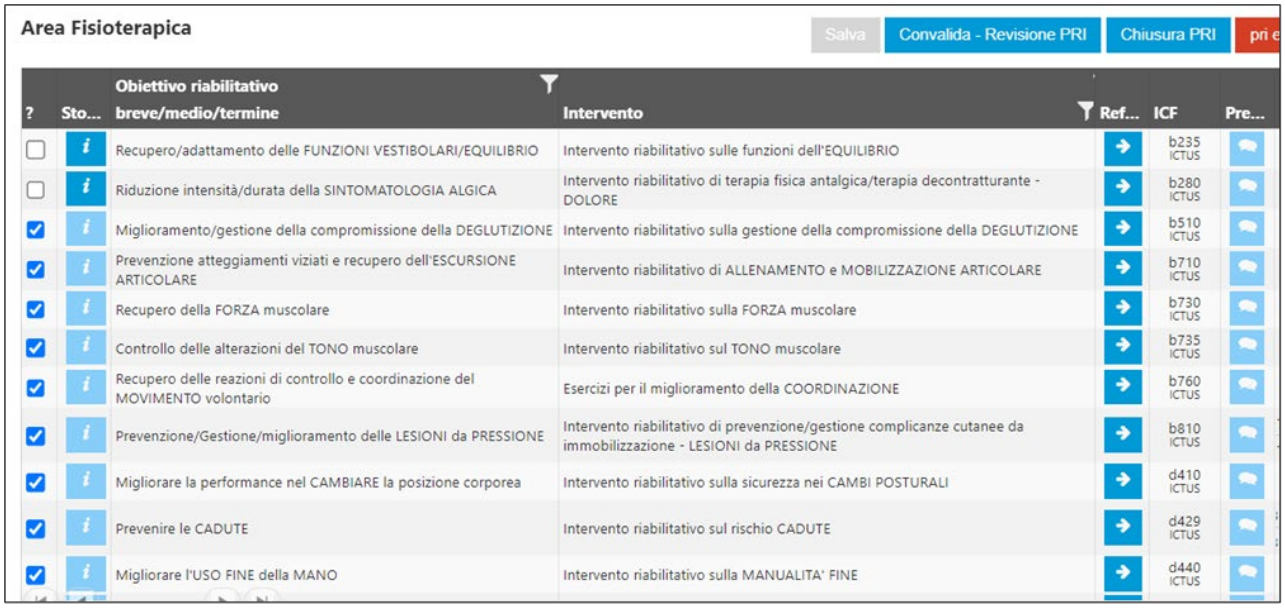


Figure 16 Definition of the Rehabilitation objects and activities for the physiotherapist

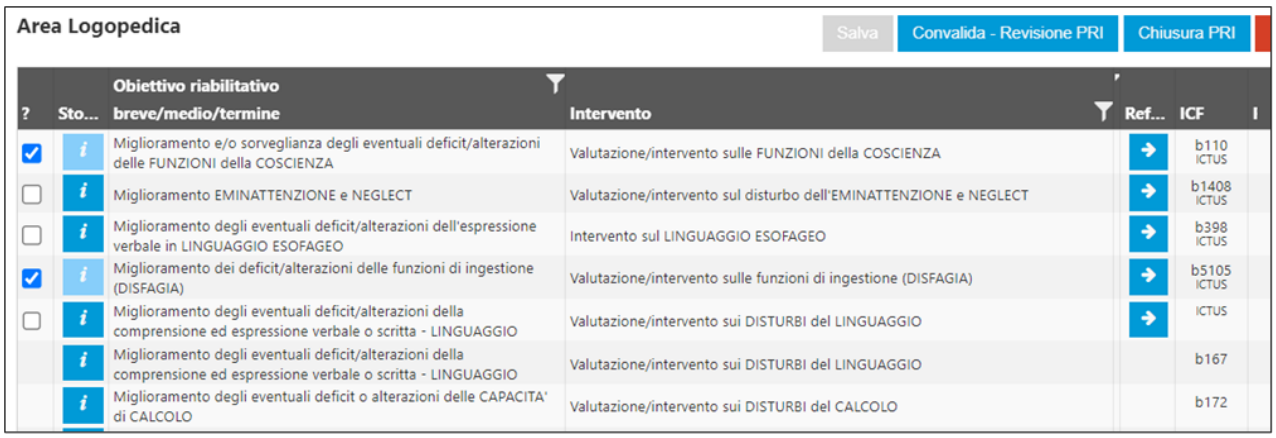


Figure 17 Definition of the Rehabilitation objects and activities for the speech therapist

In view of a personalized medicine, during the rehabilitation process, the PRI-pri may be revised modifying rehabilitation objectives and/or involving new professionals; all the possible changes in the PRI-pri are traced both in the software and in new reports which are sent to the EHR repository.

Patient's assessment

The patient is assessed by all the rehabilitative figures involved in the process and at the end of the assessment a report is produced which converges in the HER digitalized repository.
The physiotherapist performs a completely paperless assessment thanks to an internally developed digitalized tool (Assessment Manager – finalist at the EHealth4All) that supports the evaluation with predefined evaluation protocols.

Figure 18 Assessment Manager

Rehabilitation activities
Up to now the rehabilitation team mainly organizes the patient's activities on paper and pencil instruments, whereas in some identified ICSM Institutes the physiotherapists schedule the activities in a software specifically designed to support the agenda definition, by integrating the information available on scheduled activities in A4H Agendas (Dedalus Italia spa) (Cardellini et al 2023 in press). The development of the software Q-Rehab (SurgiQ s.r.l.) and its deployment in the ICS Maugeri institutes is still in progress.

Each rehabilitation objective (defined by an ICF b or d code) is associated to internally defined rehabilitation procedures, which in turn are univocally linked to an ICD9 procedure code. The rehabilitation professional, through the software Q-Rehab, indicates every day the rehabilitation activities performed with the patient and writes a diary note, which is merged in the EHR with the patient's diary.

Speech therapist
Rehabilitation objective: b5105 rehabilitation of ingestion functions (dysphagia)
Rehabilitation activities – broad definition: assessment and intervention on the ingestion functions
Rehabilitation activities – examples of analytic definition:
Secretions assessment – ICD9 code: 9301
Dysphagia rehabilitation (swallowing) – ICD9 code: 9389
Dysphagia training – ICD9 code: 9382

Physiotherapist
Rehabilitation objective: b730 muscle strength gain
Rehabilitation activities – broad definition: rehabilitative intervention on muscle strength gain
Rehabilitation activities – examples of analytic definition:
Cycle-ergometer training – ICD9 code 9312

High technology rehabilitation (Prokin) – ICD9 code 9389

Discharge
At the end of the rehabilitation program, the patient is reassessed in order to define the rehabilitation objectives achievement and final digitalized reports are produced, according to each specific team competency.