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

FIT4MEDROB

D7.3

DATA MANAGEMENT AND DATA ANALYTICS PLATFORM

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: R. Bellazzi (UNIPV), G. Nicora (UNIPV), G. D'Addio (ICSM), G. Pagano (ICSM), F. Amitrano (ICSM), A. Coccia (ICSM), A. Malovini (ICSM), V. Tibollo (ICSM), D. Pistilli (EUS), G. Bondani (EUS), F. Rebecchi (EUS), A. Foschi (EUS)

Due date of deliverable: 31/05/2024
Actual submission date: 20/09/2024
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	31/05/2024	First release
1.1	20/09/2024	Revised executive summary following external reviewers' suggestions.

TABLE OF CONTENTS

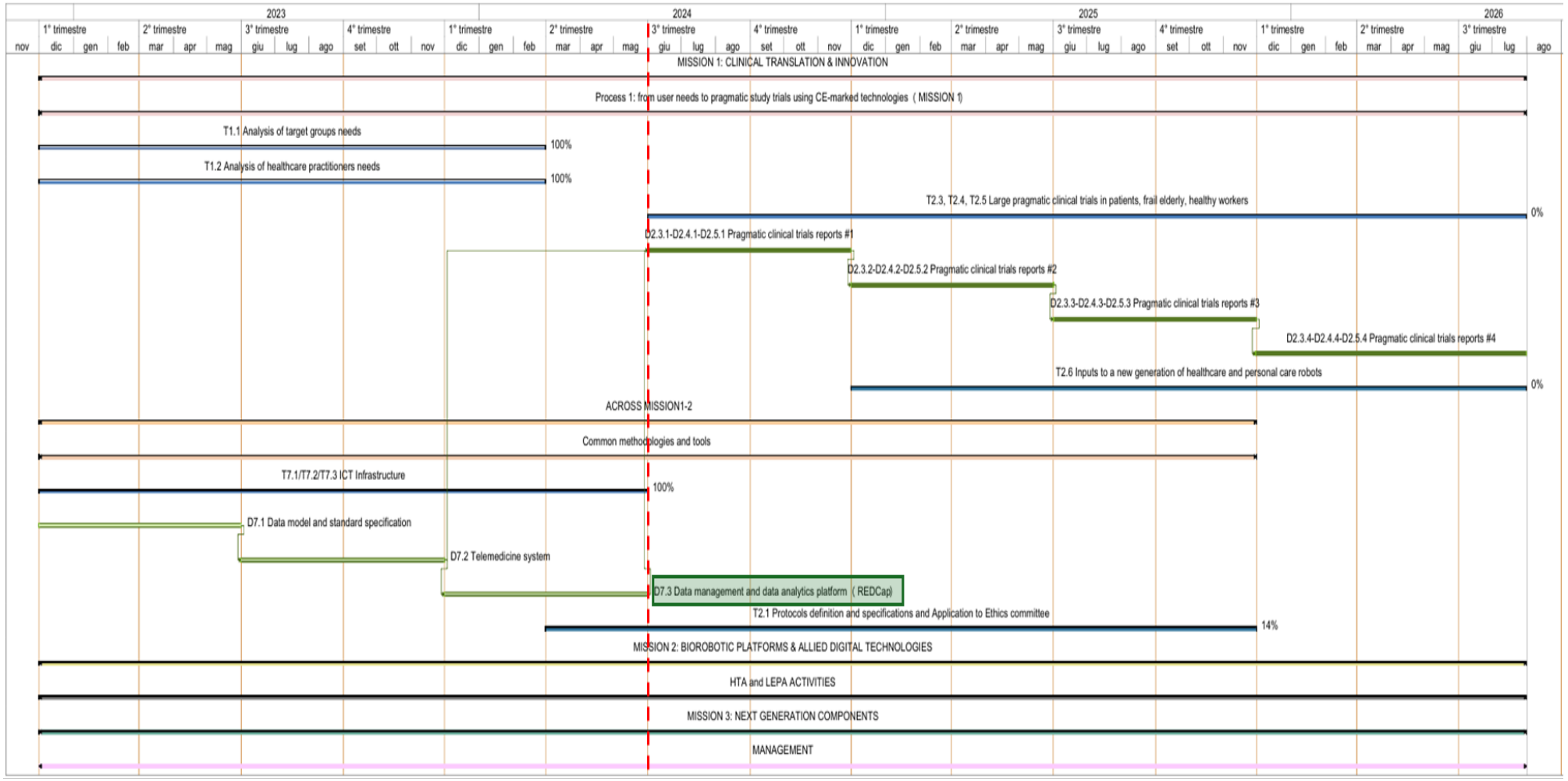
1 EXECUTIVE SUMMARY

The deliverable has been conducted after an analysis of the scenario for the implementation of ICT solutions as a set of allied technologies to medical robotics in rehabilitation. First, it describes the architecture for data collection and analytics (called in short: “monitoring tool”, or “RedCap tool”, or “prototype data platform”), implemented to support Fit4MedRob pragmatic trials (D2.1.1, month 18). The proposed architecture addresses existing gaps in rehabilitation data management, focusing on innovative solutions. The key feature is the commitment to data interoperability, adhering to standards like HL7 FHIR and proposing practical implementation strategies. The architecture, developed through multidisciplinary collaboration within the Consortium, aims to revolutionize data management in rehabilitation, addressing the integration of data generated by robot-assisted interventions, advanced analytics challenges and GDPR compliance, achieved by encryption and differential privacy.

The monitoring tool here presented will be used to **monitor all trials** envisioned in Fit4MedRob (so far: 16 large pragmatic trials plus 54 explanatory trials) and to **collect and store clinical and robot data** in the pragmatic trials (GDPR compliant). The tool is designed to provide for each trial, at a given time, the actual situation of the study (e.g. waiting for the approval from EC, enrollment, percentage of trial execution, dropped patients, etc), as well as the expected/planned situation so to obtain a direct measure of the progress.

As an example of applicability this deliverable reports an in-depth detail of the data currently under collection at multiple centres of ICS Maugeri for one of the pragmatic trials (StrokeFit4, see also D2.1.1). This case serves as an example of the information that will be jointly analyzed once the pragmatic trials of the project will be completed. Upcoming clinical trials with a large cohort of rehabilitation patients will evaluate its real-world effectiveness and impact on patient care.

It is recalled that this activity is cross-disciplinary and pertains to both Mission1 and Mission2, as the monitoring tool is pivotal for the execution of all study trials, both the pragmatic and the exploratory ones. **The completion rate of this task is 100% fully in line with the foreseen plan**, as described in the following chart. The monitoring tool will be continuously updated throughout the rest of Fit4MedRob to take into account feedback from the experimental trials and correlated practical issues that may emerge.


2023
2024
2025
2026
1° trimestre
2° trimestre
3° trimestre
4° trimestre
nov dic gen feb mar apr mag giu lug ago set ott nov dic gen feb mar apr mag giu lug ago set ott nov dic gen feb mar apr mag giu lug ago set ott nov dic gen feb mar apr mag giu lug ago
MISSION 1: CLINICAL TRANSLATION & INNOVATION
Process 1: from user needs to pragmatic study trials using CE-marked technologies (MISSION 1)
T1.1 Analysis of target groups needs
100%
T1.2 Analysis of healthcare practitioners needs
100%
T2.3, T2.4, T2.5 Large pragmatic clinical trials in patients, frail elderly, healthy workers
0%
D2.3.1-D2.4.1-D2.5.1 Pragmatic clinical trials reports #1
D2.3.2-D2.4.2-D2.5.2 Pragmatic clinical trials reports #2
D2.3.3-D2.4.3-D2.5.3 Pragmatic clinical trials reports #3
D2.3.4-D2.4.4-D2.5.4 Pragmatic clinical trials reports #4
T2.6 Inputs to a new generation of healthcare and personal care robots
0%
ACROSS MISSION1-2
Common methodologies and tools
T7.1/T7.2/T7.3 ICT Infrastructure
100%
D7.1 Data model and standard specification
D7.2 Telemedicine system
D7.3 Data management and data analytics platform (REDCap)
T2.1 Protocols definition and specifications and Application to Ethics committee
14%
MISSION 2: BIOROBOTIC PLATFORMS & ALLIED DIGITAL TECHNOLOGIES
HTA and LEPA ACTIVITIES
MISSION 3: NEXT GENERATION COMPONENTS
MANAGEMENT

2 SETTING THE SCENE: OBJECTIVES AND OVERVIEW OF CHALLENGES IN IMPLEMENTING ICT REHABILITATION ENVIRONMENT IN REAL CLINICAL PRACTICE

Fit4MedRob 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 biorobotics and allied digital technologies, and of continuum of care paradigms. Such approaches take advantage of ICT technologies in all phases of the rehabilitation process, from prevention up to home care in the chronic phase. In this context, it is important to identify the potentialities, the challenges, and the possible barriers in the introduction of interventions based on innovative technologies in the real rehabilitation practice. It is important to understand what data will be realistically collected by ICT tools and devices that will be integrated in the data analytics platforms.

Healthcare systems need to address and respond to the demographic and epidemiological trends of an aging population that is living longer compared to average life expectancy of the previous decades [1]. Longer life expectancy often implies higher prevalence of chronic diseases and related disabilities [2]. In this context, rehabilitation is proposed as a key health strategy to manage future healthcare demands. The goal of a rehabilitation program is to enhance patients' functioning ability, ensuring a safe and stable transition phase in their recovery toward community reintegration. Rehabilitation programs usually include a specific training period focusing on activities of daily living within a simulated home and community environment before discharge.

Rehabilitation can occur in various settings, including hospitals and homes, each offering unique benefits tailored to the patient's needs. In hospital settings, rehabilitation programs often involve intensive, multidisciplinary care delivered by a team of healthcare professionals, including physiotherapists and occupational therapists. These programs are well-suited for individuals requiring acute care following surgeries, injuries, or significant medical events, with access to specialized equipment and around-the-clock medical supervision. On the other hand, home-based rehabilitation provides a more personalized approach, focusing on promoting independence and functional recovery within the familiar environment of the patient's home. This setting is particularly beneficial for individuals with chronic conditions, mobility limitations, or those transitioning from hospital care to home. Home-based rehabilitation allows focusing interventions on daily activities, thanks to programs involving tailored exercise regimens, adaptive equipment recommendations, and caregiver education, empowering patients to actively participate in their recovery while maintaining their daily routines and social connections.

The choice of rehabilitation setting depends on factors such as the patient's medical condition, functional status, social support network, and personal preferences. Both hospital and home-based rehabilitation play integral roles in optimizing outcomes and enhancing the overall well-being of individuals undergoing rehabilitation. Recent research results have shown that home-based rehabilitation seems to be as effective as other forms of rehabilitation [3] [4].

Since 2005, the World Health Organization has urged member states, "to develop the infrastructure for information and communication technologies (ICT) for health as deemed appropriate to promote equitable, affordable, and universal access to their benefits, and to continue to work with information and telecommunication agencies and other partners in order to reduce costs and make eHealth successful" [5]. When it comes to disability, the Convention on the Rights of Persons with Disabilities (United Nations General Assembly, 2006) expressly highlights the potential benefits of ICT to promote independence and participation of people with disabilities. ICT has been proven to have positive effects in terms of providing alternative ways to deliver rehabilitation services. ICT presents new possibilities that can increase the quality of the rehabilitation process, by allowing for easier access to services and treatment outcomes [6]. There are several key features of ICT that underline its relevance in the rehabilitation field.

First, ICT plays an important role in patients' remote monitoring and tele-rehabilitation. In fact, it enables healthcare professionals to monitor patients' progress remotely and to provide them with real-time feedback, thus extending rehabilitation beyond traditional clinical settings. Tele-rehabilitation platforms allow for personalized exercise programs, virtual consultations, and remote supervision. These features make rehabilitation more convenient and accessible, particularly for individuals with mobility limitations, living in rural areas and/or for elderly people. Real-time patient's remote supervision allows subjects to feel more confident about the correct execution of exercises and to increase the level of motivation, thus improving the participation in everyday life with the possibility to involve relatives in the rehabilitation process.

ICT permits the collection and analysis of vast amounts of patient data, enabling healthcare providers to tailor rehabilitation programs based on individual needs and preferences. Thanks to ICT, data collection becomes more efficient and scalable, as automated processes can be employed to gather large volumes of data across diverse geographical locations. ICT facilitates real-time data collection through various digital devices and sensors, allowing organizations to gather information promptly and accurately. This capability enhances decision-making processes by providing up-to-date insights into the patients' condition and into the evolution of the rehabilitative interventions. ICT supports the integration of heterogeneous information, including structured and unstructured data from various sources such as social media, IoT devices, and online platforms. Data collected can be categorized according to their characteristics:

- demographic information
- administrative data (e.g. admission, discharge and transfer data)
- clinical data (e.g., diagnoses and results from laboratory exams),
- functional status, quality of life and physical assessment
- data related to the exercise performance in the rehabilitation program activities.

ICT is particularly promising in collecting patient-reported information (ePROMs), enabling healthcare providers to periodically receive data collections directly from the patients' perspective.

ICT also facilitates personalized interventions exploiting advanced technologies such as virtual reality, augmented reality, and gamification, to enhance rehabilitation exercises by creating immersive and engaging environments. These interactive tools not only motivate patients to participate actively in the rehabilitation activities but also to facilitate motor learning, cognitive training, and functional retraining, particularly in neuro- and musculoskeletal rehabilitation.

Biosignals and instrumental data derived from the robotic intervention (such as range of motion, force, position of center of mass) play a crucial role in rehabilitation in hospital settings because they can provide objective data about the patient's physiological state. This information helps clinicians monitor the progress of rehabilitation more accurately than relying solely on subjective reports from the patient and data from clinical scales administered by therapists. For instance, electromyography can assess muscle activity and provide insights into muscular function and recovery, while electroencephalography can support the assessment of brain activity and complement the evaluation of the cognitive function. ICT is particularly suitable to detect and analyze biosignals, providing objective assessments, guiding treatment plans, and allowing for an early detection of complications. Outside the hospital, wearable sensors offer the opportunity to monitor a variety of physiological parameters (movement, heart rate, sleep, temperature, blood glucose concentration, blood pressure, oxygen saturation) and are nowadays affordable and accessible to the public. Although these devices are not able to provide the same performances characterizing instruments commonly available in hospital settings, they offer the advantage of continuous and non-invasive monitoring of clinically relevant parameters. However, the sheer volume of commercially available sensors and applications has further highlighted the challenge of software integration, the problem of storing large amounts of data and the issue of providing meaningful indicators able to effectively summarize the health condition of the patient.

Finally, ICT can provide solutions to enhance communication and collaboration among multidisciplinary rehabilitation teams. Such characteristics allow for care coordination and interdisciplinary consultations through secure messaging systems, video conferencing and electronic platforms to exchange information, discuss treatment plans, ensure continuity of care and finally leading to more integrated rehabilitation services.

The implementation of an effective ICT-based rehabilitation framework in the clinical practice should encompass all the phases of the rehabilitation intervention (see Deliverable 7.1), consisting of:

1. prescription of the intervention;
2. delivery of the intervention;
3. monitoring and assessment of the intervention.

During the execution of the intervention, the system must be able to gather data related to the single activity performed by the patient and to the overall rehabilitation plan to assess the progress of the patient.

The adoption of ICT solutions poses increasing demands on healthcare organizations and professionals regarding how to deliver rehabilitation services. In fact, despite the potentials and the advantages of ICT solutions, their use in real clinical settings is still limited [7]. ICT solutions are usually provided by public and private suppliers. A multisource environment can supply hospitals with flexible, innovative, and cost-effective solutions needed to navigate the complexities of modern healthcare delivery, thus helping in cost savings and service improvements. On the other hand, integrating multiple vendors' solutions in the context of in-hospital rehabilitation can be challenging due to several reasons, as discussed in the following.

One of the primary challenges is ensuring that the different systems from various vendors can communicate effectively with each other and with the hospital information system. Often, these systems use different protocols, data formats, and standards, making seamless integration difficult. Interoperability issues are the result of the fragmentation of service delivery that complicates the task of integration and governance, essential to deliver effective ICT services. As mentioned above, each vendor may have its own data standards and structures. Harmonizing these disparate data formats into a unified format for analysis and decision-making can be complex and time-consuming. For example, electronic health records (EHRs) may follow proprietary formats or different standards such as HL7-FHIR [8][9]. Even when standards exist, their implementation may vary across different systems and vendors. For instance, regarding biosignals, there is no gold standard available to enable consistent acquisition, storage, presentation, analysis and sharing of data, even if several projects are addressing the issue [10]. This lack of standardization can lead to inconsistencies in data interpretation, interoperability failures, and increased complexity in systems integration efforts. Data generated by different systems within the same hospital are often siloed, meaning that they are stored in separate databases or repositories. This fragmentation inhibits data sharing and integration, limiting the ability to gather knowledge from comprehensive patient records. Integrating multiple vendor solutions implies a significant cost increase for the hospital, also because ICT solutions require maintenance, and possibly customization and data migration. However, it is not recommended to rely on a single proprietary solution because healthcare providers could be locked in that ecosystem. This aspect could limit hospitals' flexibility in switching to alternative providers in the future to choose the best solutions for their specific needs, thus limiting the innovation in the long term.

Another crucial aspect is the integration of the above-mentioned multi-vendor ICT solutions into the hospital information system (HIS), which is designed mainly with clinical, administrative, and reporting purposes. In most cases, HIS can be outdated and can rely on older technical solutions that might not be compatible with the introduction of new or updated technologies.

Besides technical aspects, *data security* and *data quality* must be considered as well. Integrating multiple systems increases the attack surface for potential security breaches and unauthorized access, posing at risk patient data privacy. The integration of multiple vendors' solutions requires rigorous testing to ensure that all the components can work together, without compromising patient safety and data integrity. Policies to govern data sharing, patient consents, and interoperability standards must be defined and regulated. Complying with the legal framework adds complexity to the implementation efforts as well. In the last part of the deliverable, we will show a case study that exemplifies the steps needed to share data among institutions within the Fit4MedRob pragmatic trials.

At the organizational level, readiness to use ICT can impact the extent to which new solutions are integrated into practice. Managing organizational changes is critical to the success of ICT multi-vendor integration initiatives. Hospital leadership must effectively communicate the benefits of integration, address staff concerns, and facilitate a smooth transition to the new system. The hospital staff consists of heterogeneous professional figures, including rehabilitation specialists, nurses, engineers and administrative personnel. Moreover, it is fundamental to include also patients and caregivers. Each category needs adequate training and support to fully exploit the ICT system. Providing training programs and ongoing support is essential but it demands intense resource use. Implementing new systems and integrating them into existing workflows can disrupt hospital operations, leading to potential inefficiencies and resistance from staff members who need to adapt to new processes. To minimize the negative impact related to the change of established routines, the new solutions should be accurately customized to meet specific needs, while remaining scalable to permit future growth. Customization and scalability are two key characteristics for the successful adoption of ICT solutions in the rehabilitation field, preventing the non-acceptance of the developed service by its end-users. Another approach is to involve the end-users early and continuously during the design process, starting from requirement analysis. This user-centered design approach situates the final user as the cornerstone of the research and development process since the success or failure of a technological solution depends on user's acceptance [11]. However, this kind of approach is common in ad hoc and project-based solutions, but it is less frequent in commercial companies, having potentially different goals and expectations compared to medical professionals.

In literature, the final users' perspective is a key factor to consider for the adoption of ICT solutions in the real clinical practice. ICT solutions evolve fast and keeping pace with novel and sophisticated rehabilitation programs is difficult for both healthcare professionals and patients. Their adherence to the use of ICT rehabilitation solutions is related to their digital health literacy, defined as the "the ability to search, find, understand, evaluate health information from electronic sources and apply the knowledge gained to address or solve a health problem" [12]. ICT solutions are used by the same clinicians and health professionals who practice traditional medicine, and their rapid evolution is placing an increasing demand for digital literacy among these clinical experts [13]. There have been increasing initiatives to fill this gap, for

example by establishing new medical disciplines [14], able to incorporate digital health as part of the physician's curriculum. Besides personal skills, there are other factors that can influence the adoption of ICT technologies. In literature [15], factors like collaboration, leadership and individual and team characteristics are considered in addition to individual features such as age, experience, attitude, and knowledge. Collaboration, in particular the creation of multidisciplinary teams, is considered the preferred method to deliver ICT service [16].
Within Fit4MedRob, we have identified data to be collected following the functional workflow reported below:

1. Patient's initial assessment and rehabilitation plan: based on patient's initial assessment, the clinician prepares the individual rehabilitation plan that contains the description of the rehabilitation objectives and the specification of the intervention, (i.e. the temporal progression and intensity of the exercises to be performed within the period of rehabilitation activity).
2. Delivery plan: the clinician selects the delivery setting (hospital, nursing home, patient's home) and defines the final version of the rehabilitation plan in accordance with the initial plan (point 1). According to the selected settings, the plan can be transferred directly to the patient (for example, through smartphone apps) or to the clinical specialists involved inside the hospital or nursing homes (for example, through dedicated dashboards).
3. Execution plan: the patient executes the rehabilitation activity, managed, and assisted by the clinicians. During the activity, the patient can interact with robotic devices and can be equipped with sensors able to monitor the execution of the exercise (e.g. 3D cameras, inertial sensors). All the devices must be connected in a virtual environment. At the end of the activity, proper indicators must be used to gather feedback from the patients, for example to evaluate exercise tolerance or the easiness of the execution.
4. Monitoring plan: the descriptive data of each exercise session (including instrumental data from devices, clinical scales and feedback from patients) are sent to a repository. The transmission can be asynchronous or synchronous according to the characteristics of the rehabilitation intervention. The data collected are visualized in a monitoring platform available to all the specialists involved in the patient's treatment. According to the monitoring results the intervention needs to be adapted by iterating steps 3 and 4 and, if needed, redesigning the overall intervention plan.

As reported in Deliverable 7.1, to implement the described workflow, the Fit4MedRob project proposes an ICT infrastructure that includes several main components: a centralized control system, robust networking, sensor networks, cloud computing, and mobile devices.

- The first component is a centralized control system provided with a dashboard to manage and control the various robotics interventions of the rehabilitation activity. This system should allow for the definition and the remote monitoring (in case of telerehabilitation intervention) and for the remote access and control of the robot or sensors used. Authorized personnel should be able to monitor the operational settings from a central location via a dedicated dashboard. The system could also embed a Digital Support System (DSS), which can monitor patient's exercise motion and assess patient's performance using sensors and machine learning algorithms to generate quantitative measurements for clinical assessment.
- A robust networking system is essential for exchanging data and instructions between the centralized control system and the devices. A high speed, reliable network is necessary to ensure the device can operate efficiently and without interruption. The network should also keep into account security issues, implementing security measures such as strong encryption, authentication, proper firewalls settings, intrusion prevention and detection systems. Regular updates and maintenance of the software used is essential to guarantee the efficacy of the security measure. This aspect is particularly critical in the healthcare settings because it protects patient confidentiality and reduces the risk of data breaches. A robust cybersecurity strategy also enables healthcare organizations to comply with relevant regulations. The system should also be based on a standardized data communication protocol to collect sensing data and transmit them in an interoperable format.
- According to the delivery setting chosen for the rehabilitation plan, sensors should be placed in strategic locations to provide real-time data on the rehabilitation process and on the activities of the robots used. The sensors can include cameras, microphones, temperature and pressure sensors, physical activity sensors, like wristbands. To allow for remote assistance, the patient should be equipped with a chatbot (supplied via mobile app for example) to receive instructions and reminders and to ask questions in natural language about the rehabilitation activities. The mobile app can be used also to inform the patients about the activity plan and the progress recorded in a simple and concise way, in order to overcome, as much as possible, issues related with usability.

- To store and process the data generated during the rehabilitation activities, the architecture should comprehend a computing environment. Fit4MedRob proposes a Cloud computing architecture based on a European infrastructure able to provide real-time analytics and insights which can be used to optimize the performance of the devices and improve patient outcomes. Cloud computing can provide a scalable and cost-effective solution to store and process data about environment, patient activity, and device performance. The healthcare organizations can store large amounts of data without investing in expensive on-site storage solutions that are difficult to quantify in advance in terms of extension. Fit4MedRob exploits the concept of data lake, a centralized repository that allows for the storage of structured and unstructured data. Data lakes allow to store data in its native format, and this facilitates a large range of use cases and improves data reusability, especially when compared to the schema-on-write approach applied in data warehouses, where data is transformed prior to the actual storage to fit a predefined schema [17].

Data lakes bring several advantages, by:

- permitting the integration of data from disparate sources;
- accommodating large volumes of data;
- enabling advanced analytics and data processing techniques.

Data lakes can be also shared across healthcare organizations, enabling collaborations between healthcare providers. However, storing such massive amounts of raw data leads to new challenges, spanning from the general data modelling, and indexing for concise querying to the integration of suitable and scalable compute capabilities. Beside this, data lakes require robust data governance practices, including policies and procedures for data access, data quality, data privacy and data security.

Overall, addressing the challenges of integrating ICT rehabilitation interventions in real clinical settings demands strategic planning, stakeholder collaboration and ongoing assessment. Overcoming barriers to interoperability, considering resource limitations and data integrity concerns, necessitates a comprehensive approach, involving stakeholders from healthcare providers to policymakers. By prioritizing open standards, investing in interoperable solutions, and fostering a culture of collaboration, the full potential of ICT in healthcare can be realized, leading to improved delivery and patient outcomes. To achieve a successful implementation, it is necessary to invest not only on technological proficiency but also on addressing human and organizational factors, enhancing multidisciplinary team experiences. To this end, Fit4MedRob proposes an interdisciplinary intervention including physicians to raise clinical questions, bioengineers to find technical solutions and social scientists to ensure the solutions can be actually adopted. The project aims to cover technological, economical, legal and policy gaps currently present in the Italian healthcare system, that have prevented nationwide clinical adoption of ICT-based rehabilitation interventions by creating a technical solution able to solve most of the issues here analyzed.

Within the Fit4MedRob project, it clearly emerged the need of implementing an overall architecture for collecting the data of pragmatic trials in a centralized way. Thanks to the deep analysis described in this section, it has been possible to design and implement an overall infrastructure that is oriented to data analytics, that will be reported in the following of the deliverable.

3 METHODS

The intersection of the healthcare sector and data science holds considerable transformative potential, particularly within the domain of rehabilitation where patient outcomes can significantly benefit from personalized interventions and predictive analysis [18]. As the volume and complexity of healthcare data grows, the need for sophisticated analytics platforms indeed becomes fundamental to harness these resources effectively. Considering this context, we have designed a prototype healthcare data platform, specifically tailored for the rehabilitation sector. The prototype platform is envisioned as a supportive tool designed to streamline the analysis of the vast data generated within rehabilitation settings, thereby facilitating improved clinical decisions and outcomes, as clearly stated in the previous section. Utilizing a micro-services architecture and leveraging open-source technologies, the prototype aims to set a benchmark for interoperability, scalability, and practical utility in clinical trials and research. This work is at the basis of the data analytics platform of Fit4MedRob [19],[20],[21]. Through this development, in particular, this project seeks to underscore the importance of targeted analytics and data interoperability in the healthcare sector and pave the way for future advancements in rehabilitation technologies in the frame of the Fit4MedRob initiative.

Healthcare data analytics has evolved from basic descriptive analytics to more complex predictive and prescriptive analytics, leveraging vast amounts of data to inform clinical decision-making and policy [22]. In rehabilitation, the application of analytics promises to tailor interventions to individual patient needs, optimize treatment outcomes, and enhance operational efficiencies [23]. Despite these potentials, the adoption of advanced analytics in rehabilitation has been relatively slow, hindered by challenges in data interoperability, privacy concerns, and the complexity of rehabilitation data [24], [25].

Rehabilitation environments are characterized by diverse data types, including clinical assessments, patient-reported outcomes, sensor data from medical devices, and more, as fully described in the previous section [26]. This diversity necessitates a robust analytics platform capable of integrating and interpreting heterogeneous data sources to support personalized rehabilitation pathways [27]. Moreover, the dynamic nature of rehabilitation, where patient conditions and treatment responses continually evolve, calls for real-time analytics to adjust treatment plans promptly [23], [28].

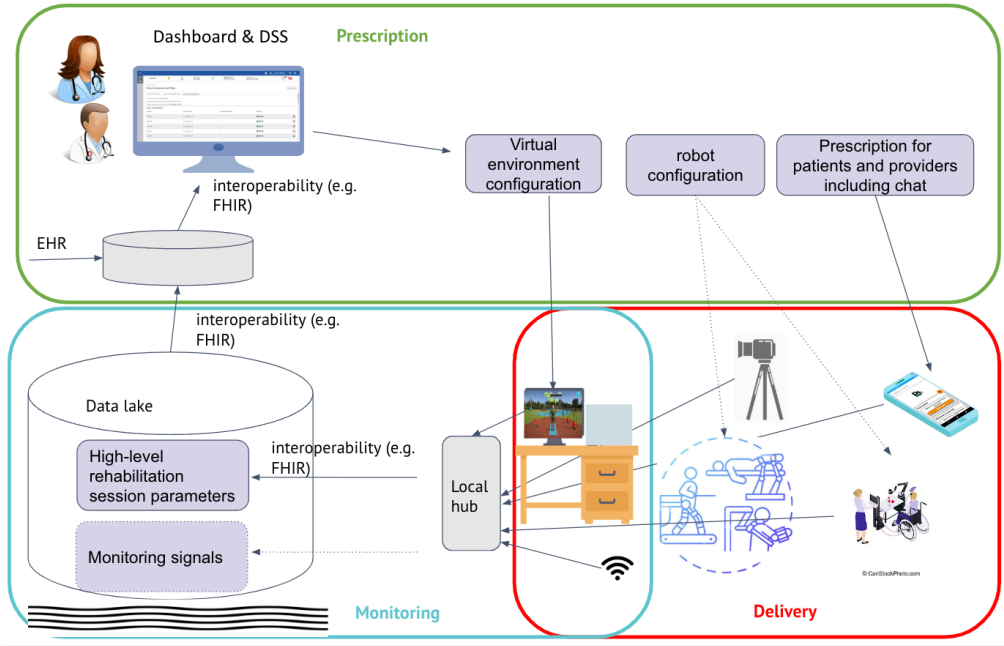


Fig. 1 General ICT architecture.

Deliverable D7.1 has defined a general ICT architecture to support robotic rehabilitation, reported in Figure 1. This architecture has been adapted to the context of supporting the Fit4MedRob clinical trials, which have the specific need of collecting clinical information, instrumental session data and biosignals, and to provide tools for the analysis required by the different trials.

Therefore, the components of the general architecture discussed in this deliverable are: the monitoring dashboard and DSS components as well as the data lake (left part of the figure from top to bottom). In the following we will collectively refer to these two components as the data analytics platform. The design of the components is general enough to allow for collecting data coming from different devices, as it will be needed by the different trials foreseen in the project.

In the design of the data analytics platform, we considered several fundamental elements.

- *Data interoperability.* Data interoperability in healthcare is crucial for the seamless exchange and utilization of healthcare information across different systems and stakeholders. Key standards facilitating this include Health Level Seven (HL7), Fast Healthcare Interoperability Resources (FHIR) [29], and the Digital Imaging and Communications in Medicine (DICOM) [30]. HL7 and FHIR are instrumental in promoting efficient data exchange and integration, with FHIR being noted for its adaptability and straightforward implementation, supporting a broad spectrum of healthcare applications, including those in rehabilitation [32]. DICOM is essential for managing medical images, ensuring compatibility across imaging devices and healthcare systems, which is crucial in rehabilitation for treatment planning and progress monitoring [32].
- *Privacy and security.* The privacy and security of patient data are paramount, especially with the increasing reliance on big data analytics. Anonymization techniques such as k-anonymity, l-diversity, and differential privacy have been explored to protect patient confidentiality while allowing valuable insights to be gleaned from healthcare data [33].

However, the effectiveness of these techniques in maintaining data utility while ensuring privacy in complex rehabilitation data landscapes remains an area of ongoing research [34].

- *Data augmentation.* Parallel to anonymization, machine learning (ML) models for data augmentation present innovative solutions to address the dual challenges of data scarcity and privacy. Data augmentation techniques, pivotal in enhancing ML model performance, are adept at artificially expanding dataset sizes without necessitating additional real samples. This is particularly valuable in healthcare, where acquiring extensive datasets is often hampered by privacy regulations and the sensitive nature of patient data. Techniques such as synthetic data generation, as discussed by [35], and advanced deep learning approaches for medical imaging reviewed by [36], exemplify the strides being made in this domain. Moreover, the integration of Generative Adversarial Networks (GANs) as highlighted in [37] and the utilization of frameworks like MONAI, which [38] and [39] delve into, demonstrate the breadth of innovation in applying ML to augment healthcare data while maintaining integrity and privacy.
- *Cloud computing.* Cloud computing has emerged as a powerful tool for managing and analyzing healthcare data, offering scalability, flexibility, and cost-efficiency. Several healthcare organizations [40],[41],[42],[43] have adopted cloud services to facilitate data storage, integration, and analytics [44]. The use of cloud services in rehabilitation analytics promises to enhance data accessibility and collaboration among multidisciplinary teams, driving improvements in patient care and outcomes.

4 RESULTS

- Referring to the general architecture, the data analytics platform is conceived with a multi-layered architecture, designed to seamlessly integrate, process, and analyze diverse data types inherent to rehabilitation environments, including both unstructured data (such as images) and structured data (like tables). At its core, the platform comprises six distinct layers (Figure 2), each serving a crucial function in the data lifecycle.

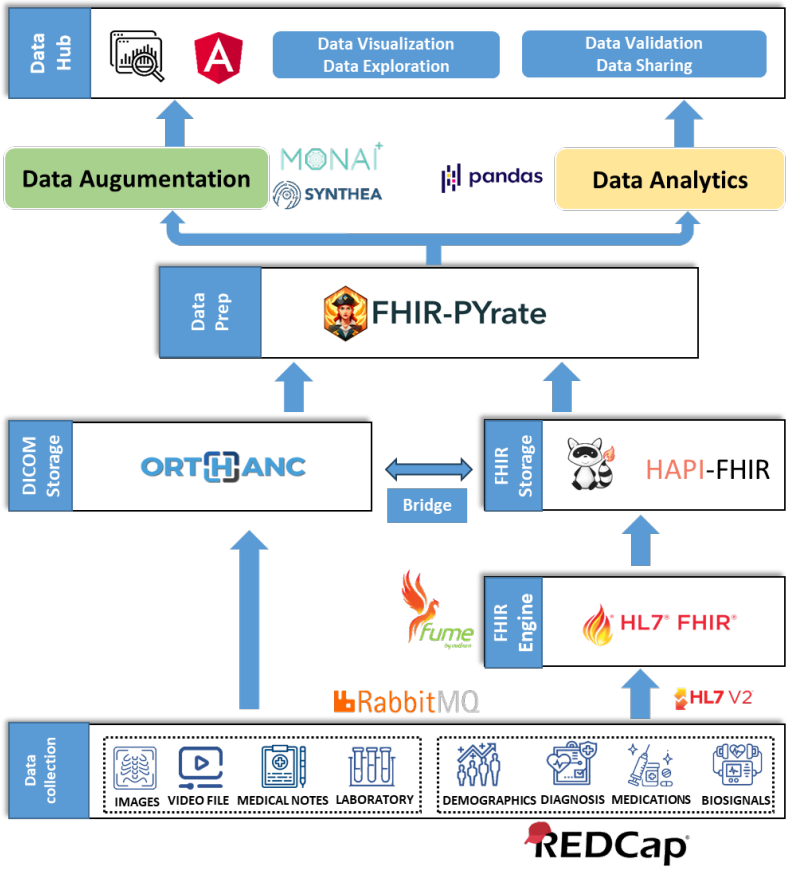


Fig. 2 The Data Analytics Platform architecture.

Data Collection Layer: this foundational layer is tasked with aggregating data from various sources, including medical devices, wearable sensors, and electronic health records. Utilizing REST API protocols, it ensures a smooth and secure data flow into the data analytics platform, facilitating data acquisition for rehabilitation processes. Within this layer, special recognition is given to the REDCap (Research Electronic Data Capture) data capture tool [45], [46], identified as the main source for obtaining clinical information of patients involved in rehabilitation programs. REDCap is a web-based application from the REDCap Consortium that consents to gather secure and reliable data for research studies. The link between the Fit4MedRob REDCap instance (described at chapter 4) and the data analytics platform is formed via the API services provided by the application, allowing for the comprehensive transfer of all stored patient data. Through the data collection layer, the data analytics is also able to interface with the most common rehabilitation devices such as motion analysis or gait analysis devices, and to query these to acquire the rehabilitative session data of a particular patient. In the initial phase of prototype testing, the devices with which the data analytics platform will be tested will be those of Tecnobody s.r.l. (a member of the Fit4MedRob consortium), after which the approach will also be extended to devices from other vendors. In the initial phase, there will also be a specific focus on structured data and the processing of HL7 message streams. Subsequently, the experimentation will also be extended to sources of unstructured information.

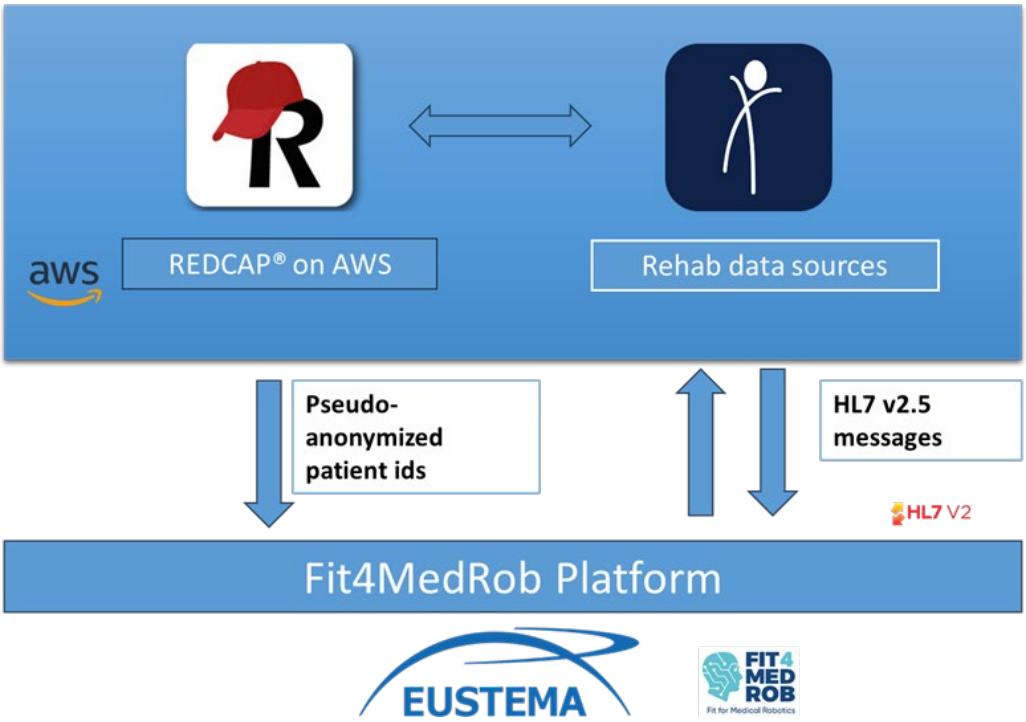


Fig. 3 Data collection process.

Data Integration Layer: this layer is equipped with a specialized engine designed to convert incoming data to the FHIR format, leveraging the open-source Fume FHIR converter library [47]. Unstructured data (i.e., medical images) do not undergo any conversion because they are already in the standardized DICOM format, which is natively supported by medical devices. The Data Integration Layer ensures data standardization across the board, enabling universal compatibility and easing its assimilation into the extensive healthcare data ecosystem. Furthermore, it enhances data integrity and facilitates smoother data integration. The conversion of incoming data, such as HL7 v2.5 messages from robotic rehabilitation devices, into FHIR JSON format, is facilitated by the Fume FHIR converter library. This tool uses custom mapping files to dictate how data elements in the input messages are translated into FHIR resources. The converter engine reads the input data, applies the mapping rules defined in the custom mapping files, and generates a JSON-formatted output that conforms to the FHIR standard. This includes creating JSON objects for each FHIR resource type involved, such as Patient, Encounter, or Observation, that are ready to be integrated into the broader healthcare

data ecosystem. This data is standardized and interoperable, making it easier to share across different systems and platforms.

The mapping files are configuration manifest that specify how data from the input messages should be mapped to corresponding FHIR resources; they are written in the Fume custom programming language, FLASH. FLASH is an enhancement designed by the Fume authors to the FHIR shorthand (FSH) language and allows any FSH assignment rule to accept a dynamic value expressed as a FUME expression. In fact, the Fume converter can connect to a FHIR server so it can check and define the ID field to assign to a specific resource (the resource Identifier) and then manage possible cross-references with other resources already ingested. The mapping file can also define necessary transformations or conversions.

For example, the mapping file for any HL7 v2.5 messages conversion outlines which elements in the HL7 message correspond to specific attributes in the FHIR JSON structure and the appropriate field conversions. For instance, date formats need to be changed from YYYYMMDDhhmmss (common in HL7) to YYYY-MM-DDThh:mm:ss+zz:zz (used in FHIR). An illustrative example shown in the figure below demonstrates how an HL7 v2.5 synthetic message is transformed. The message is composed of several segments and details a “Equilibrio Monopodalico Destro” test performed on the Tecnobody s.r.l. PROKIN 252 robot, mapping it to an Encounter resource in FHIR v4.0.1, with associated observations mapped to Observation resources.

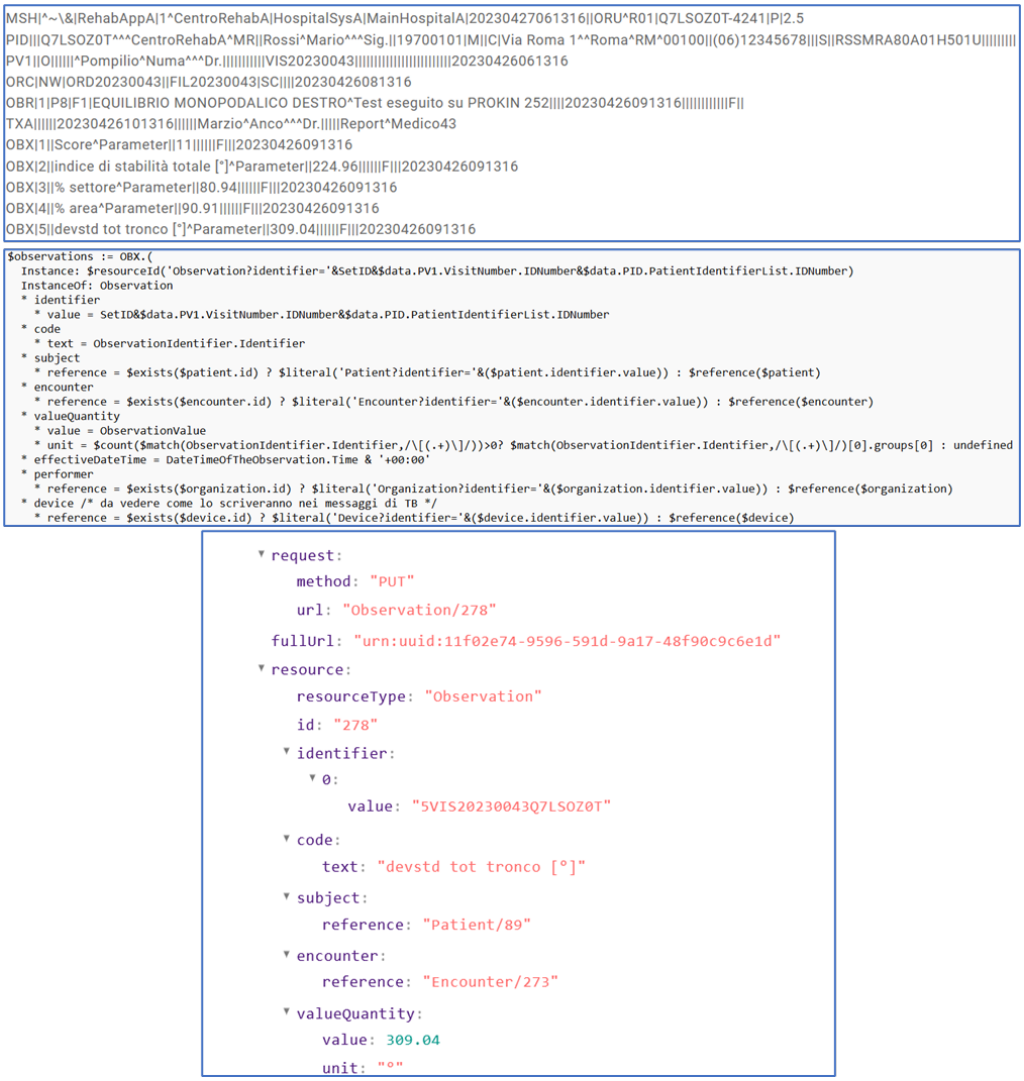


Fig. 4 A synthetic incoming HL7 v2.5 message generated by a Tecnobody s.r.l. rehabilitation robot; a portion of the corresponding Fume mapping file; a portion of the output FHIR JSON ready to be ingested in the FHIR server.

Data Storage Layer: in this layer, data is organized within a scalable and easily accessible data lake. It utilizes the HAPI-FHIR [48] framework to manage structured data in alignment with the FHIR standard and incorporates the open-source PACS system Orthanc [49] for handling unstructured data such as DICOM images. This setup not only ensures compliance with healthcare data interoperability standards but also supports synchronous data ingestion. As depicted in the referenced diagram, the Orthanc service receives DICOM resources which are synchronized with the structured data management executed by the HAPI-FHIR service. This ensures a continuous and efficient data flow within the system. A key component of this synchronization is the "synchronizer," a Python-based tool that aligns DICOM metadata between the HAPI-FHIR and Orthanc services. This includes ensuring that each DICOM file in Orthanc corresponds to an "ImagingStudy" resource in the HAPI-FHIR server. This synchronizer is tailored specifically to manage metadata related to DICOM resources, such as study, series, and instance identifiers, aligning them between the structured data in HAPI-FHIR and the unstructured data in Orthanc. By doing so, it guarantees a seamless flow of information between the structured and unstructured domains, enhancing data integrity and retrieval efficiency across the healthcare system. The synchronizer operates continuously, updating changes between the HAPI-FHIR server and Orthanc, ensuring that metadata for any DICOM study is consistent across both platforms. The synchronizer will be tested in the second phase of the prototype experimentation, when data acquisition will also include bioimaging devices.

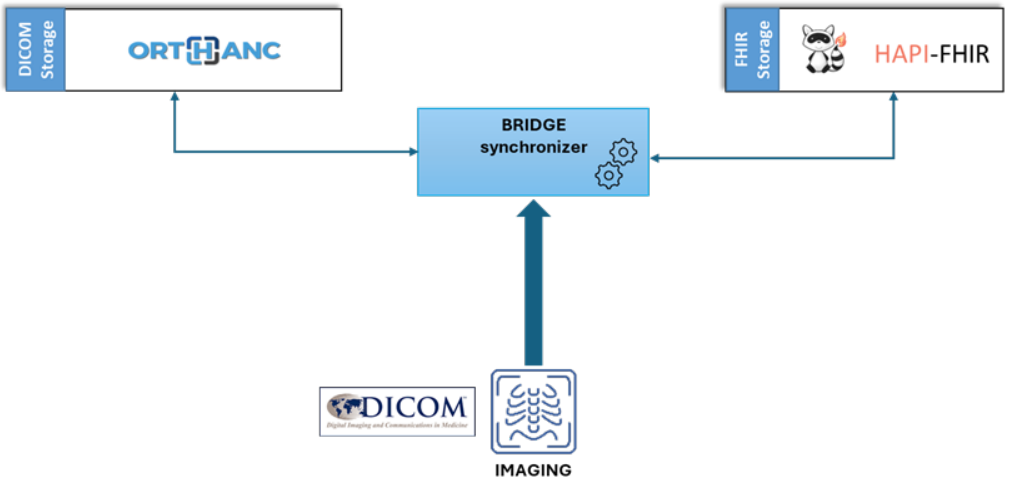


Fig. 5 PACS Synchronizer.

In the case of the structured data, no synchronization is necessary, because HAPI-FHIR is the only storage support used to store them.

Data Preparation Layer: Preparing the data for analysis, this layer focuses on transforming raw data into a format suitable for visualization and further analysis. Its logic is based on FHIR-PYrate library [50]. FHIR-PYrate is a Python package designed to simplify querying FHIR servers. Specifically, it facilitates efficient data retrieval and provides a user-friendly interface for customizing queries and extracting data as pandas DataFrames. Additionally, FHIR-PYrate includes a DICOMDownloader class, automating the download of DICOM studies from PACS. This functionality enables the collection of imaging data referenced in FHIR resources and supports structured data export.

Data Augmentation Layer: To support advanced analytics and machine learning models, this layer applies AI techniques to generate synthetic datasets. These datasets are crucial for overcoming data scarcity issues, enhancing model training, and supporting predictive analytics, thereby facilitating the development of personalized rehabilitation strategies. This layer is implemented with Synthea [51] and MONAI library. Synthea is an open-source software tool that simulates realistic but synthetic patient data based on configurable models of diseases, treatments, and healthcare interactions. It is designed to produce comprehensive datasets that include detailed medical histories for synthetic patients, ensuring that the data covers a wide range of healthcare scenarios and conditions. MONAI library is instead a PyTorch-based framework specifically tailored for healthcare imaging. It provides developers with tools and pre-built workflows for designing and deploying AI models that are robust and reproducible. The library supports a range of tasks crucial for

medical imaging such as segmentation, classification, and registration, and is optimized for high performance with extensive support for various imaging modalities.

Data Analytics Layer: This layer is a pivotal component of the platform, designed to leverage advanced data analysis methods, exploiting both predictive and prescriptive models to support diagnostic and therapeutic processes within the rehabilitation domain. The development of these tools, which will include AI models, will utilize data collected during the project's experimental phase, as well as supplementary datasets obtained from open-access sources in the literature. The datasets will be appropriately enhanced through the Data Augmentation Layer described previously. Key functions will include:

- Predictive Analytics: models to forecast patient recovery and identify potential complications early.
- Prescriptive Analytics: models that recommend personalized treatment plans based on data analysis.

These features will support advanced, data-driven decision-making in rehabilitation, to improve patient outcomes and clinical efficiency. This layer will be built on popular ML python packages like scikit-learn, PyTorch or NLTK and it will employ both traditional ML methods (i.e. regression models, random forests, etc..) and deep learning models. The application of these machine learning techniques will be guided by rigorous validation and testing protocols to ensure their reliability and effectiveness in clinical settings, and it will be preceded by a careful phase of data analysis and data preparation.

Data Visualization Layer: This layer represents the final stage in the data flow, offering user-friendly interfaces to interact with the functionalities offered by the previously outlined backend layers. It is built as a single-page web application utilizing Angular v17 as its development framework. For visualizing data through graphs and tables, it employs various JavaScript libraries. The web application ensures secure interactions via token-based, stateless session management with the server and delivers customized features to users according to their authorization levels. The platform envisions two primary user profiles: the "admin" profile for clinical center administrators and the "user" profile for individual therapists and clinicians. Admin users have the ability to create and manage user accounts within their own rehabilitation center. If necessary, additional user types can be identified during the validation phase based on new combinations of functionality access needs and authorization grants.

The features provided by this layer include:

- *Data ingestion view*: in this view, the user is provided with a set of features for data ingestion from the REDCap server using a three-step wizard: 1) initiation of the ingestion process with the option to select which data to acquire; 2) display of the status of the acquisition process message by message; 3) display of the status of the acquired messages with the option to view both the original HL7 message and the version converted to FHIR (Figure 6).
- *Data search view*: in this view, the user can search for previously acquired data through tabs dedicated to the type of data, with specific filters, and display the data in paginated and sortable tables. Here, the user can select the records of interest for visualization through charts of various types (Figure 7).
- *Data visualization view*: after the selection of data to visualize in the previous view, here the user can make plots of various types, from line charts to bar charts and others, and can choose which information to plot on the x-axis. This view also provides a basic calculation of statistical information, such as mean and standard deviation based on the data selection (Figure 8).
- *User management view*: here, the administrator user can manage other users present in the system, modify existing ones, and create new ones according to needs.

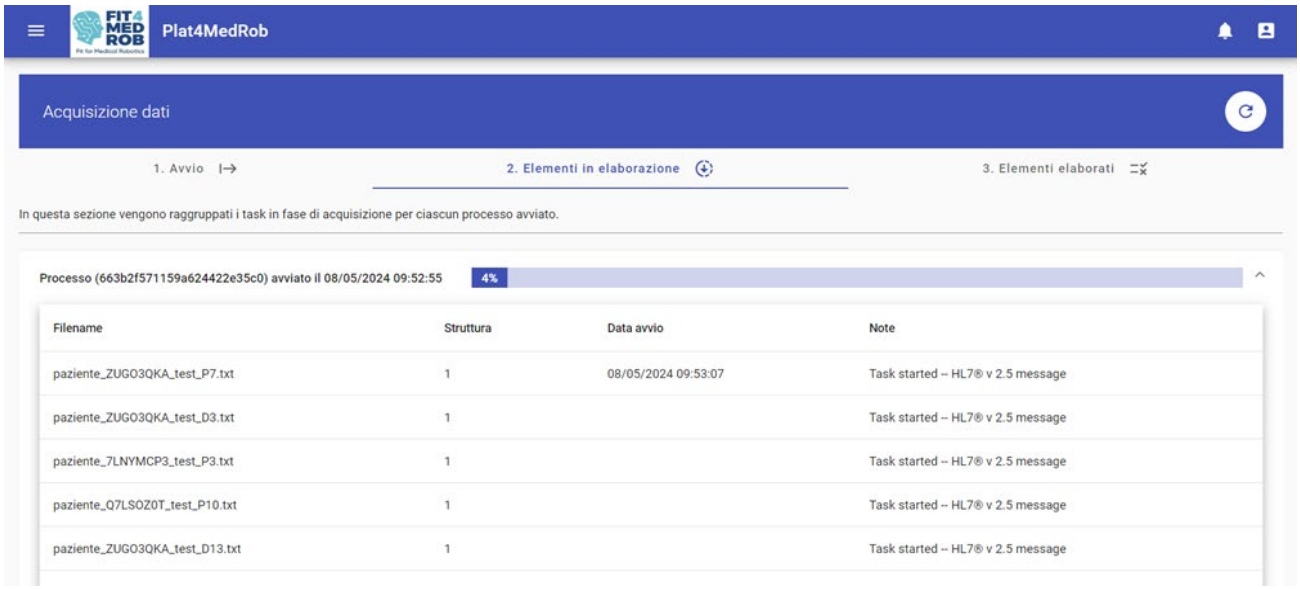


Fig. 6 Data ingestion view, display of the status of the acquisition process message by message.

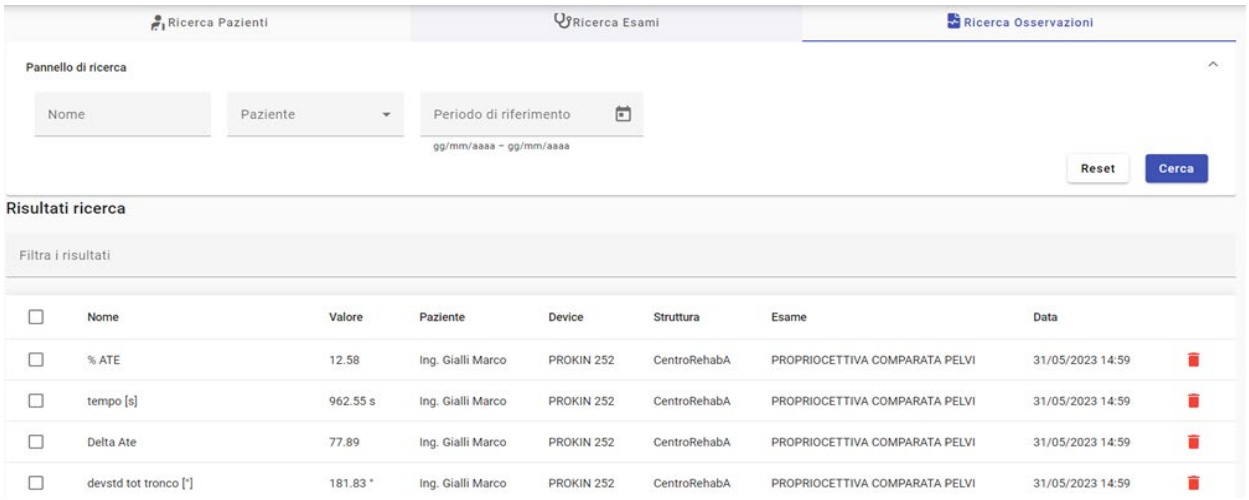


Fig. 7 Data search view, displaying the results of an observations search.

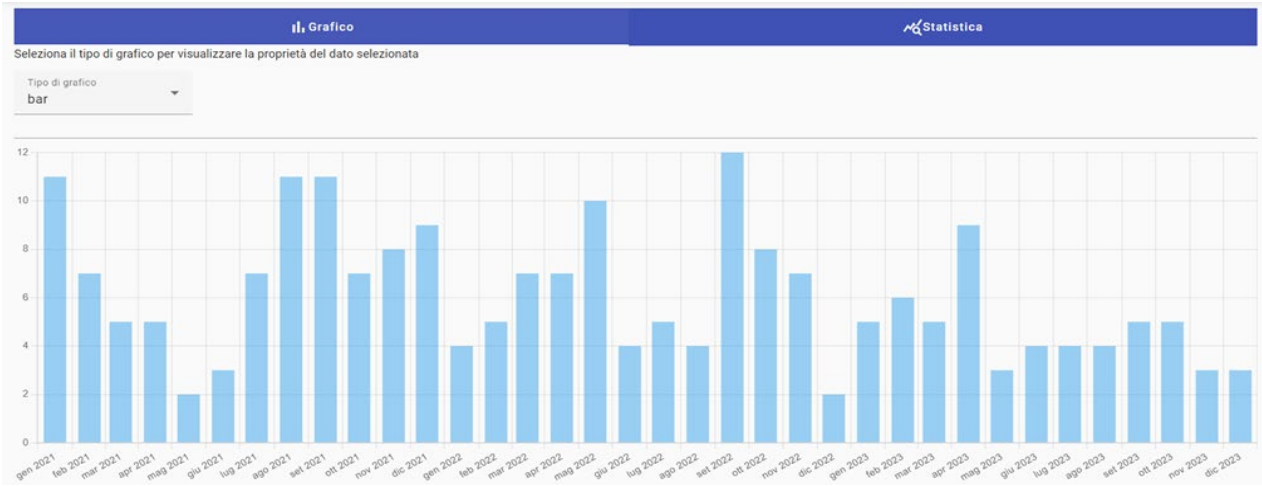


Fig. 8 Data analytics view, an example of a bar chart displaying the number of encounters in a range of time.

The implementation of the individual microservices that make up the architecture was carried out using Python 3.9 and the FastAPI library [52]. All components are containerized via Docker and designed as autonomous building blocks. Additionally, the processing of incoming structured data is managed through asynchronous queues with Python library Celery [53] and a message broker. For storing information about asynchronous tasks, a MongoDB instance is employed. Within the Fit4MedRob initiative, the prototype testing plan of the data analytics platform is structured into five main phases, which will be described below. The prototype testing will involve only a subset of rehabilitation centers affiliated with the Fit4MedRob consortium and will primarily focus on analyzing the effectiveness of digital and robotic technologies in clinical rehabilitation practice. It will have a duration of approximately two years and will conclude with the end of the Fit4MedRob project. It is also important to emphasize that this testing plan will be executed concurrently with the Fit4MedRob pragmatic trials, which will involve the rehabilitation of a sample of patients recruited nationwide. During the prototype development plan, continuous collaboration with consortium partners, including Tecnobody s.r.l., University of Pavia, University of Genoa, University of Pisa, ICS Maugeri, Fondazione Don Gnocchi, and many others will be of crucial value. These partnerships will provide specific expertise, clinical insights, and access to rehabilitation environments necessary for the success of the data analytics platform.

Phase 1: Requirement Analysis and Initial Design (Q4 2023)
This initial phase, which took place in the last four months of the past year, primarily focused on the detailed collection of the platform's functional requirements from all consortium partners, including rehabilitation centers, technology providers, and research institutes. In this first step, the project team's task was to identify the specific needs of end users, data security standards, and interoperability specifications that the platform must meet.

Phase 2: Design and Development (Q1-Q2 2024)
Leveraging the information acquired in Phase 1, the project team will proceed with the design and development of the prototype. This phase involves the architectural design of the data analytics platform, selection and integration of open-source components, and development of the data collection, storage, preparation, augmentation, analytics and visualization layers. Preliminary testing ensured the platform's basic functionality and readiness for further integration and testing within the clinical trial setting. In this phase we have already implemented and deployed the Fit4MedRob REDCap instance, described in the following section.

Phase 3: Integration and Comprehensive Testing (Q2 2024)
This stage, which directly precedes the pilot trial, will focus on integrating the healthcare data platform with the clinical trial infrastructure, ensuring seamless data flow and compatibility. Comprehensive testing will cover functionality, security, and interoperability within the trial's technological ecosystem, especially in relation to the REDCap server and the rehabilitation robots. In the first months of testing, experimentation will primarily involve structured data coming from rehab devices. After this phase, bioimaging and biosignals will also be included. During this phase, the initial AI models will be designed and trained using the collected structured data. The training process will leverage both collected data and augmented datasets to ensure robustness and accuracy.

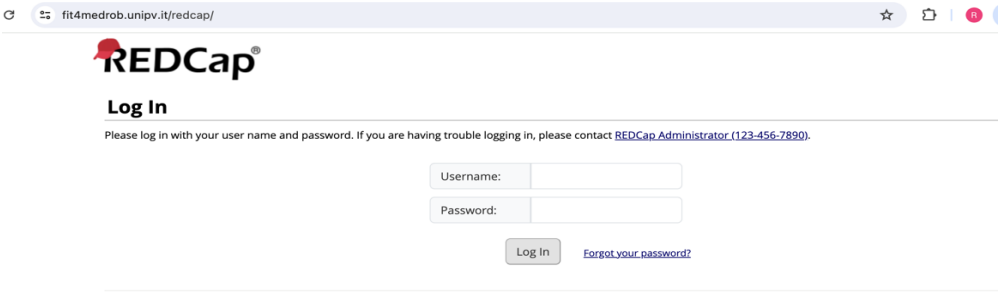
Phase 4: Deployment and Evaluation within the Clinical Trials (Q3 2024 - Q2 2026)
The prototype platform will be set up in a managed environment at the involved rehabilitation centers. This phase aims to evaluate the platform's effectiveness in real-world settings, gathering data on its impact on clinical decision-making, patient outcomes, and operational efficiencies. Feedback from clinicians, researchers, and patients will be critical to identifying areas for refinement. Feedback will be obtained with dedicated interviews, focus groups and validated usability questionnaires. AI models will be continuously refined and tested during this phase based on real-world data and feedback from users, ensuring they meet the practical needs of clinical environments.

Phase 5: Final Assessment (Q2 2026 onwards)
At the conclusion of the study, a comprehensive final evaluation will be conducted, considering the feedback and data collected during the project. This assessment will focus on the impact of the platform within the study and will evaluate its scalability for broader application in rehabilitation settings (involving rehabilitation centers that were not included in the initial experimentation phase). Furthermore, this phase will explore potential refinements and future developments to enhance the functionality and utility of the platform.

During the study and the experimentation, the ongoing collaboration with clinical groups and Fit4MedRob consortium partners will ensure that the platform is always aligned with the project goals and the therapists' activities. Regular feedback cycles will also facilitate continuous improvement and the adaptation of the platform to the evolving needs of the study and its participants.

5 THE FIT4MEDROB REDCAP INSTANCE

We have installed and deployed a REDCap instance at the UNIPV servers hosted on Amazon cloud. The platform can be reached at https://fit4medrob.unipv.it/redcap/.



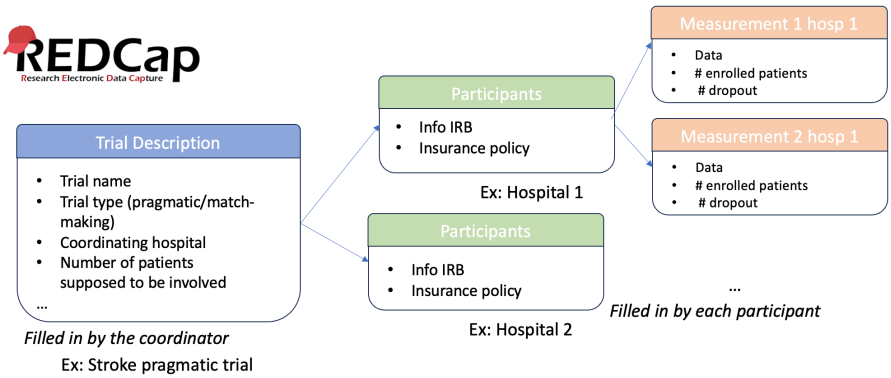
The first two projects that will be deployed are: i) a clinical trial monitoring dashboard, which will collect general data regarding the status of all Fit4MedRob trials; ii) the Stroke pragmatic trial clinical data collection project.

5.1 TRIAL MONITORING DASHBOARD

The Monitoring dashboard will serve as a tool to track the various phases of Fit4MedRob clinical trials. More specifically, for each trial, a new record will be added within the dashboard, containing the following information (see Figure 10a below):

1. Trial description (type of the trial, coordinating center, number of patients to be enrolled, its status, for instance if enrollment has started or if IRB approval is still undergoing).
2. Participants information: a record for each hospital participating in the trial (if multi-centric). Information reported will be related to IRB and policy assurance.
3. Once the trial has started, each participant will periodically measure relevant information, such as the number of patients enrolled, the number of dropouts, the number of patients that completed the study.

The dashboard will show the current status of Fit4MedRob clinical trials as in the example depicted in Figure 10b.



(a)

Record ID	Descrizione trial	Partecipanti	Istituti Clinici Scientifici Maugeri - Pavia	FDG - Salerno	COT - Messina	FDG - Milano	FDG - Roma	FDG - Firenze	FPUCBM	Gaslini - Genova	Inail	Istituti Clinici Scientifici Maugeri - Bari	Istituti Clinici Scientifici Maugeri - Milano	Istituti Clinici Scientifici Maugeri - Montescano	Istituti Clinici Scientifici Maugeri - Tradate	Mondino - Pavia	San Martino - Genova	Stella Maris
1 Trial pragmatico stroke																		
2 Parkinson trial																		

(b)

Fig. 10 (a) Outline of the information required to create a new Record; (b) Dashboard with Fit4MedRob clinical trials.

An example of the Trial description form is reported in the following page:

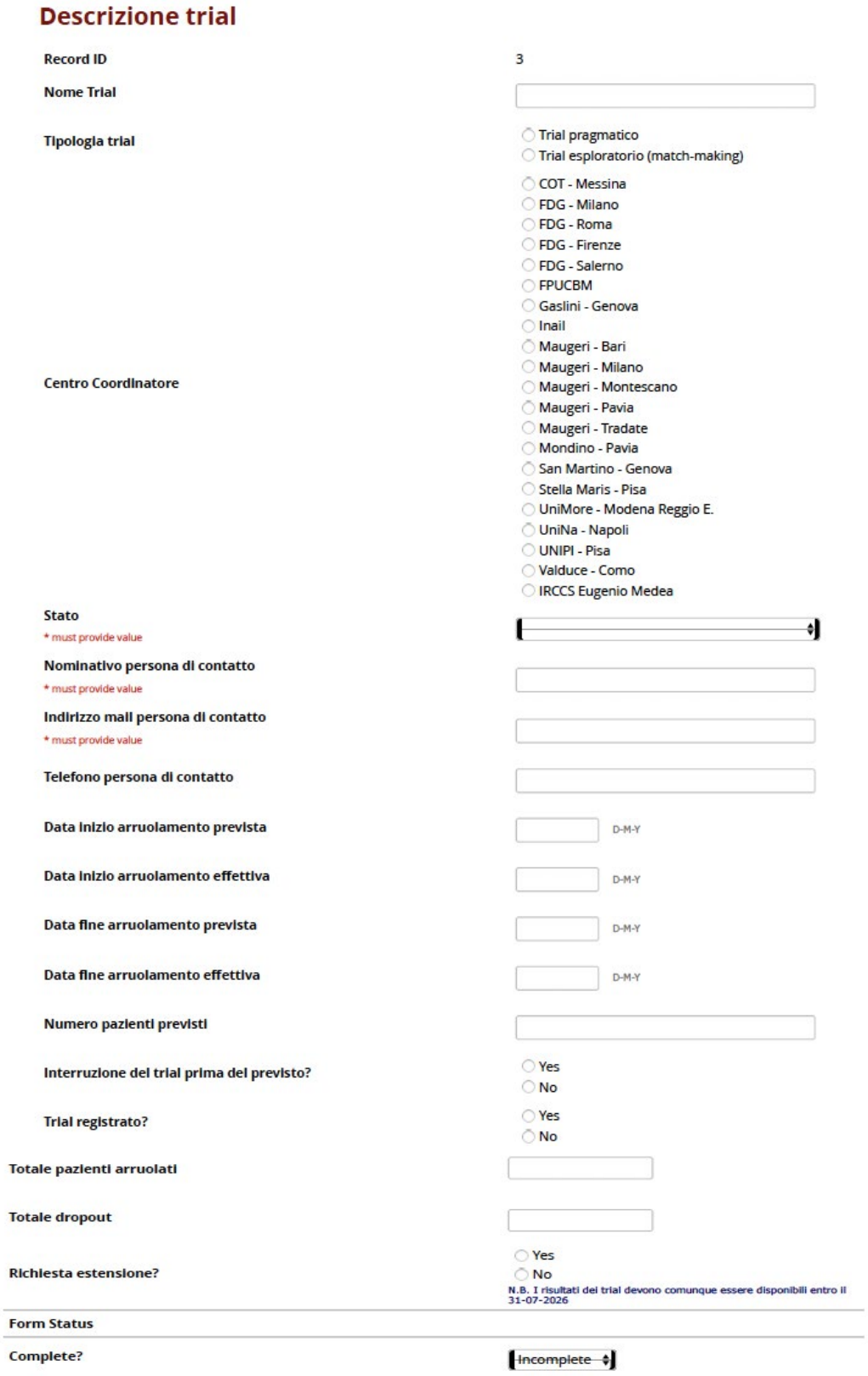


Fig. 11 Trial description form.

5.2 STROKE PRAGMATIC TRIAL

Within the project, a multi-center pragmatic trial will be performed to compare robotics rehabilitation with standard rehabilitation for stroke patients. We have developed a prototype of the REDCap project that will be exploited to collect data during the trial (Figure 12). Different data will be collected during the various phases of the trial:

- During enrollment and hospitalization, the patient's history, comorbidities, and general information, such as age, gender, BMI, will be collected. Relevant clinical scale, such as the Modified Ranking scale, will be reported before starting the rehabilitative treatment.
- During the follow-up of the trial, clinical scales for measuring the patient's outcome will be collected.

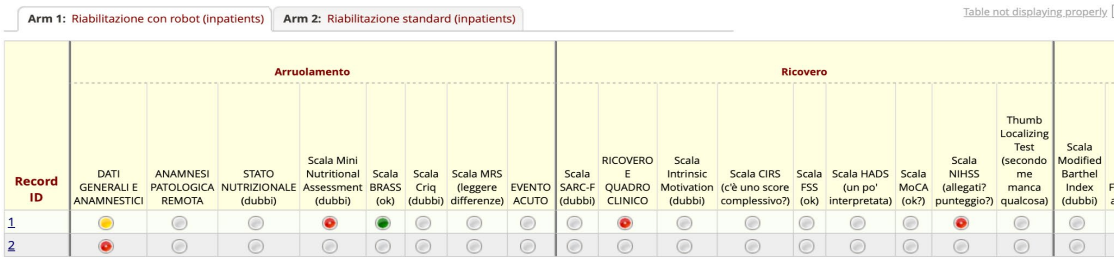


Fig. 12 Example of the REDCap prototyoe developed to collect data during trial.

An example of data collection during enrollment is shown in Figure 13 reported below:

Fig. 13 Example of data collection during enrolment.

6 FEEDING THE DATA ANALYTICS INFRASTRUCTURE

To provide a clear vision of the requirements of the data analytics platform from all the pragmatic trials, we have detailed the different types of data that will be collected by three equipment that will cover a wide range of pragmatic trials within Fit4MedRob: Treadmill WalkerView, Prokin static and dynamic stabilometric platforms by Tecnobody and Hocoma's Erigo system.

A. Treadmill WalkerView

Walker View is a compact and complete motion analysis and biomechanics laboratory that combines the functionality of walking and running tests and corrective training.
Walker View performs Gait Analysis on patients and verifies the parameters of dynamic posture and support during walking with objective data. It provides real-time visual and acoustic feedback. The following table reports the parameters collected by Walker View along different treatments protocols.

TEST	Indexes	Parameters
Fast start	W1	Mean value Flex-ext trunk [°], ROM Flex-ext trunk [°], min flex-ext trunk [°], max flex-ext trunk [°] ROM Hip SX Flex-ext [°], min Hip SX Flex-ext [°], max Hip SX Flex-ext [°] ROM Hip DX Flex-ext [°], min Hip DX Flex-ext [°], max Hip DX Flex-ext [°] ROM Knee SX Flex-ext [°], min Knee SX Flex-ext [°], max Knee SX Flex-ext [°] ROM Knee DX Flex-ext [°], min Knee DX Flex-ext [°], max Knee DX Flex-ext [°] Duration [hh:mm:ss], distance [km], mean speed [km/h], max speed [km/h], contact time SX [s], contact time DX [s], symmetric charge [%], height difference, calories [kcal] , ROM COG mean [cm], ROM COG mean % [%]
GAIT ANALYSIS	W2	Mean value Flex-ext trunk [°], ROM Flex-ext trunk [°], min flex-ext trunk [°], max flex-ext trunk [°] ROM Hip SX Flex-ext [°], min Hip SX Flex-ext [°], max Hip SX Flex-ext [°] ROM Hip DX Flex-ext [°], min Hip DX Flex-ext [°], max Hip DX Flex-ext [°] ROM Knee SX Flex-ext [°], min Knee SX Flex-ext [°], max Knee SX Flex-ext [°] ROM Knee DX Flex-ext [°], min Knee DX Flex-ext [°], max Knee DX Flex-ext [°] Duration [hh:mm:ss], distance [km], mean speed [km/h], max speed [km/h] Vertical oscillation COG [cm - %], cadence [cycles/s], stride length SX-DX [cm], stride length (coefficient of variation) SX-DX [%], contact time SX-RH [s], symmetrical load [%]
RUN ANALYSIS	W3	Mean Flex-ext trunk [°], ROM Flex-ext trunk [°], min flex-ext trunk [°], max flex-ext trunk [°] Mean flex-lat trunk [°], ROM flex-lat trunk [°], min flex-lat trunk [°], max flex-lat trunk [°] ROM Hip SX Flex-ext [°], min Hip SX Flex-ext [°], max Hip SX Flex-ext [°] ROM Hip DX Flex-ext [°], min Hip DX Flex-ext [°], max Hip DX Flex-ext [°] ROM Knee SX Flex-ext [°], min Knee SX Flex-ext [°], max Knee SX Flex-ext [°] ROM Knee DX Flex-ext [°], min Knee DX Flex-ext [°], max Knee DX Flex-ext [°] Duration [hh:mm:ss], distance [km], average speed [km/h] Vertical oscillation COG [cm - %], cadence [cycles/s], stride length LH-RH [cm], step length (coefficient of variation) LH-RH [%], contact time LH-RH [s], symmetrical load [%]

COOPER TEST	W4	Mean Flex-ext trunk [°], ROM Flex-ext trunk [°], min flex-ext trunk [°], max flex-ext trunk [°] Mean flex-lat trunk [°], ROM flex-lat trunk [°], min flex-lat trunk [°], max flex-lat trunk [°] ROM Hip SX Flex-ext [°], min Hip SX Flex-ext [°], max Hip SX Flex-ext [°] ROM Hip DX Flex-ext [°], min Hip DX Flex-ext [°], max Hip DX Flex-ext [°] ROM Knee SX Flex-ext [°], min Knee SX Flex-ext [°], max Knee SX Flex-ext [°] ROM Knee DX Flex-ext [°], min Knee DX Flex-ext [°], max Knee DX Flex-ext [°] Duration [hh:mm:ss], distance [km], average speed [km/h] LH Pitch Length [cm], RH Pitch Length [cm], LH Contact Time [s], RH Contact Time [s], Load Symmetrical [%], Elevation gain, calories consumed [kcal], ROM COG average [cm], ROM COG average % [%], VO2 max [mL/(KG*min)], outcome [Poor-very poor- good - very good]
Protocol MINUTES WALK TEST	W5	Mean Flex-ext trunk [°], ROM Flex-ext trunk [°], min flex-ext trunk [°], max flex-ext trunk [°] Mean flex-lat trunk [°], ROM flex-lat trunk [°], min flex-lat trunk [°], max flex-lat trunk [°] ROM Hip SX Flex-ext [°], min Hip SX Flex-ext [°], max Hip SX Flex-ext [°] ROM Hip DX Flex-ext [°], min Hip DX Flex-ext [°], max Hip DX Flex-ext [°] ROM Knee SX Flex-ext [°], min Knee SX Flex-ext [°], max Knee SX Flex-ext [°] ROM Knee DX Flex-ext [°], min Knee DX Flex-ext [°], max Knee DX Flex-ext [°] Duration [hh:mm:ss], distance [km], average speed [km/h] LH Pitch Length [cm], RH Pitch Length [cm], LH Contact Time [s], RH Contact Time [s], Load Symmetrical [%], Elevation gain, calories [kcal], ROM Average COG [cm], ROM Average COG % [%] , Number of stops [int] , Result [Poor-very poor- good - very good]
BALKE TEST	W6	Mean Flex-ext trunk [°], ROM Flex-ext trunk [°], min flex-ext trunk [°], max flex-ext trunk [°] Mean flex-lat trunk [°], ROM flex-lat trunk [°], min flex-lat trunk [°], max flex-lat trunk [°] ROM Hip SX Flex-ext [°], min Hip SX Flex-ext [°], max Hip SX Flex-ext [°] ROM Hip DX Flex-ext [°], min Hip DX Flex-ext [°], max Hip DX Flex-ext [°] ROM Knee SX Flex-ext [°], min Knee SX Flex-ext [°], max Knee SX Flex-ext [°] ROM Knee DX Flex-ext [°], min Knee DX Flex-ext [°], max Knee DX Flex-ext [°] Duration [hh:mm:ss], distance [km], average speed [km/h] LH Pitch Length [cm], RH Pitch Length [cm], LH Contact Time [s], RH Contact Time [s], Load Symmetrical [%] Elevation gain, calories consumed [kcal], ROM COG average [cm], ROM COG average % [%], VO2 max [mL/(KG*min)], outcome [Poor-very poor- good - very good]

B. Prokin static and dynamic stabilometric platforms

ProKin is a system for proprioceptive-stabilometric assessment designed for the rehabilitation of the lower limbs. The technology applied to ProKin 252 sees a mechanical system and 50-level of electronic stability control connected to the software for complete integration and maximum gesture control. Thanks to these characteristics, ProKin 252 enables the patient to reconstruct the correct map of proprioceptive sensations.
The data acquisition card transfers the angular movements of the patient moving on the mobile platform to the PC. The joint movement data collected in a kinesthetic trace is displayed on the monitor. The static, single-leg dynamic with controlled load, double-leg dynamic and trunk control modes of use allow you to manage differentiated and functional

training for the recovery of the lower limbs. The software offers numerous games to train balance, capable of helping the patient escape and at the same time maintaining a high level of concentration during motor re-education.
The following table reports the parameters collected by Prokin along different treatments protocols.

TEST	INDEX	PARAMETERS
STABILOMETRY EYES OPEN	P1	Score , Area [mm^2] , Perimeter [mm], Devstd AP [mm], Devstd ML [mm], average COP Y [mm] , average COP X [mm], regression line angle [°] , Devstd tronco [°] , Average velocity AP [mm/s], Average velocity ML [mm/s]
BIPODALIC STABILOMETRY OA / OC	P2	Score , Area [mm^2] , Perimeter [mm], Devstd AP [mm], Devstd ML [mm], average COP Y [mm] , average COP X [mm], regression line angle [°] , Devstd tronco [°] , Area OC / Area OA , Perimeter OA / Perimeter OC , Average velocity AP [mm/s], Average velocity ML [mm/s]
RIGHT MONOPODALIC STABILOMETRY	P3	Score, Area, Perimeter, Devstd AP, Devstd ML, average COP Y, average COP X, regression line angle, Devstd tronco, Average velocity AP, Average velocity ML
LEFT MONOPODALIC STABILOMETRY	P4	Score, Area, Perimeter, Devstd AP, Devstd ML, average COP Y, average COP X, regression line angle, Devstd tronco , Average velocity AP, Average velocity ML
COMPARED MONOPODALIC STABILOMETRY	P5	Score, Area, Perimeter, Devstd AP, Devstd ML, average COP Y, average COP X, regression line angle, Devstd tronco, Average velocity AP, Average velocity ML
LOS	P6	Score, Devstd trunk [°] % total objectives, % side
BIPODALIC EQUILIBRIUM	P7	Score, total stability index [°], stability index AP [°], stability index ML [°], platform stability level, % sector, % area, devstd tot trunk [°]
RIGHT MONOPODALIC BALANCE	P8	Score, total stability index [°], % sector, % area, devstd tot trunk [°]
LEFT MONOPODALIC BALANCE	P9	Score, total stability index [°], % sector, % area, devstd tot trunk [°]
COMPARATIVE MONOPODALIC BALANCE	P10	Score, total stability index [°], % sector, % area, devstd tot trunk [°]
RIGHT PROPRIOCEPTIVE	P11	Score, % ATE, Mean Sector Forces [kg, Devstd tot trunk [°], % Major Stationary Sector, Time [s]
LEFT PROPRIOCEPTIVE	P12	Score, % ATE, Mean Sector Forces [kg], Devstd tot trunk [°], % Major Stationary Sector, Time [s]
PROPRIOCEPTIVE COMPARED	P13	Score, % ATE, Delta Ate, Mean forces sectors [kg], ratio devstd tot trunk LH/RH, % Major stationary sector, time [s]
BESS	P14	Oscillation index [°]
MCTSIB	P15	Oscillation index [°]
STAR TEST	P16	Score, % per area reached
COMPARATIVE PELVIC PROPRIOCEPTION	P17	Score, % ATE, Delta Ate, devstd tot trunk [°] , % Major Parking Sector, Time [s]

Tecnobody devices have an interoperability layer that will allow sending data with HL7 V2 to the Fit4Med Platform. An example of an HL7 message is shown below.

MSH|^~\&|Tecnobody|Tecnobody|||||MDM^T02||P|2.5

PID|||~

PV1||Or

OBR|1

OBX|1|ED|||^^^Base64^{\X0D\\X0A\ "TestType": 35,\X0D\\X0A\ "Results": {\X0D\\X0A\

"Speed": 5.6,\X0D\\X0A\ "average bpm": 0.0,\X0D\\X0A\ "AP": 2.1,\X0D\\X0A\ "ML": 1.1,\X0D\\X0A\ "ROM": 45.0,\X0D\\X0A\ "51234": 42.8,\X0D\\X0A\ "61234": 67.9,\X0D\\X0A\ "71234": 63.0,\X0D\\X0A\ "LP SX": 78.0,\X0D\\X0A\ "Load": 3.2,\X0D\\X0A\ "LP DX": 76.0,\X0D\\X0A\ "TC SX": 0.61,\X0D\\X0A\ "Cadence": 1.0,\X0D\\X0A\ "TC DX": 0.61,\X0D\\X0A\ "TRUNK flex-lat": 1.1,\X0D\\X0A\ """: 1.0,\X0D\\X0A\ "161234": 3.2,\X0D\\X0A\ "171234": 3.0,\X0D\\X0A\ "Knee": 7.5,\X0D\\X0A\ "191234": 7.0,\X0D\\X0A\ "Hip": 5.0,\X0D\\X0A\

"211234": 5.0\X0D\\X0A\ }\X0D\\X0A\}

C. ERIGO

Hocoma's Erigo system is designed to provide an integrated and advanced solution for physical rehabilitation, combining robotic mechanics and electrical stimulation. The export of training data plays a crucial role in evaluating the effectiveness of treatment, allowing healthcare professionals to monitor patients' progress and tailor therapies according to individual needs.

From this device, it is possible to export a variety of significant parameters, in *.xlsx* format, which include personal patient information, details of training sessions, and specific metrics related to physical performance and electrical stimulation. Exported data includes patient name, custom ID, date of birth, height, weight, BMI, training number, training date and duration, number of steps, and detailed parameters such as cadence, load, angle of inclination, range of motion, and, in the case of FES application, frequency and characteristics of the pulse, and amplitude of stimulation on multiple channels. The import of the data into the data analytics platform will be performed after conversion of the data into a proper FHIR resource. This activity will be performed in the next phases of the project.

For a more in-depth understanding, you can refer to the following figure, which shows the schema of the export of data in Excel format, clearly showing how it is organized and made accessible for detailed analysis.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC
1	Name	Custom ID	Birthdate	Height [cm	Weight [I	BMI	Training	Training Date	Duration	Steps	FES	Cadence	Loading	Tilt Angl	ROM L [°]	ROM R [°]	GF L [%]	GF R [%]	Pulse [µs	Ramp	Frequen	Amp 1 [m	Amp 2 [m	Amp 3 [m	Amp 4 [m	Amp 5 [m	Amp 6 [m	Amp 7 [m	Amp 8 [mA]
2	beatriz	dias	25/01/1991	161	56	22	1	06/02/2020 15:42	00:09:45	287	No	36	21	58	34	38	98	32											
3	beatriz	dias	25/01/1991	161	56	22	2	14/02/2020 09:52	00:03:43	76	No	22	16	5	25	39	100	100											
4	beatriz	dias	25/01/1991	161	56	22	3	04/02/2021 15:02	00:01:23	21	No	16	12	0	7	8	100	100											
5	beatriz	dias	25/01/1991	161	56	22	4	04/06/2021 12:05	00:08:11	420	No	58	83	91	31	31	100	100											
6	beatriz	dias	25/01/1991	161	56	22	5	14/06/2021 15:57	00:03:56	114	Yes	42	46	2	31	22	76	76	250	3	10	0	0	0	0	0	0	2	2
7	beatriz	dias	25/01/1991	161	56	22	6	01/09/2021 11:25	00:12:25	365	Yes	34	30	86	21	20	100	50											
8	beatriz	dias	25/01/1991	161	56	22	7	08/09/2021 14:37	00:13:10	342	No	37	31	85	31	36	0	92											
9	beatriz	dias	25/01/1991	161	56	22	8	23/09/2021 13:51	00:16:40	413	No	25	29	66	26	27	100	100											
10	beatriz	dias	25/01/1991	161	56	22	9	23/09/2021 14:12	00:16:10	505	No	37	45	76	29	28	100	100											
11	beatriz	dias	25/01/1991	161	56	22	10	23/09/2021 14:29	00:12:39	441	No	42	23	50	33	39	88	88											

7 DISCUSSION AND CONCLUSIONS

The prototype healthcare data platform for rehabilitation environments embodies several original contributions and advancements in the field of healthcare data analysis. Through its development and anticipated deployment, the project not only addresses existing gaps in rehabilitation data management but also lays the groundwork for forthcoming innovative solutions within the scope of the Fit4MedRob initiative.

At the heart of the platform is a commitment to data interoperability, achieved through adherence to and evaluation of current standards such as HL7 FHIR. The project contributes to the ongoing discourse on interoperability by

demonstrating practical implementation strategies and identifying areas where existing standards can be extended or refined to better support the complex data landscapes of rehabilitation environments.
Recognizing the paramount importance of patient privacy, the platform incorporates advanced data anonymization methods that strike a balance between the utility of data and the strict privacy demands, including those outlined by GDPR. Techniques like differential privacy and encryption stand as critical defences against the stringent privacy laws faced by entities managing both health and non-health personal data.
A key contribution of the platform is also its data visualization and augmentation capabilities, designed to support clinical decision-making. By employing AI to generate synthetic datasets and developing intuitive visualization tools, the project enhances the ability of clinicians and researchers to derive actionable insights from complex data. This contribution is particularly significant in the context of rehabilitation, where personalized treatment plans based on comprehensive data analysis can lead to improved patient outcomes.
The following will detail more specifically the limitations and innovative aspects of the solution.
The project stands out for its pioneering approach to developing a data infrastructure capable of supporting robot-assisted rehabilitation systems. This innovation primarily enables the creation of customized rehabilitation models for patients, enriching a market offering that is still very limited today. In this regard, it is worth noting that there are already entities within the consortium (and beyond) that have developed platforms for the centralized storage of data from various types of robotic devices. However, it is also true that the goal of this project is not only to create a platform that includes data from multiple robotic devices but also to integrate other types of healthcare data, such as bioimaging and signal data, into this same platform, creating a solution that is as integrated, scalable, and AI-oriented as possible. This approach will indeed enable the creation of a unified repository for data analysis and AI model experimentation in rehabilitation and beyond. Furthermore, the project introduces an unprecedented level of data interoperability, enabling the aggregation of data from diverse rehabilitation systems across different vendors and standardizing this data into a common format, whether HL7/FHIR or another accredited standard. Another notable innovation is the incorporation of data augmentation techniques within the data platform, aimed at ensuring compliance with health data privacy regulations and facilitating the development of more robust machine learning models. Data augmentation emerges as a strategic response to the common issue of limited sample sizes in healthcare data, potentially improving the creation of datasets for AI tool experimentation in rehabilitation.
Despite these innovations, the project faces several uncertainties due to its high complexity. A primary concern is the feasibility of developing an ICT platform that can effectively acquire data from rehabilitation systems of varying technologies and harmonize this data according to a single interoperability standard. Each rehabilitation machine's unique outcome specifications and data models pose significant integration challenges, raising questions about the project's ability to achieve comprehensive interoperability. The efficacy of data augmentation and anonymization techniques also remains uncertain, particularly regarding their ability to produce realistic and high-quality data. Similarly, the effectiveness of the AI models developed for the Data Analytics Layer is uncertain due to the same challenges. This uncertainty is compounded by the complexity of healthcare data, the types of data selected for experimentation, and the quality of available datasets for model training. The quality of data used depends on both the types of data collected and the methods of data collection. The variability in data sources, including structured clinical records, unstructured medical images, and biosignals, as well as the differences in data collection protocols across various rehabilitation devices, further contributes to this complexity and uncertainty. Moreover, the project's architecture must accommodate high operational loads and ensure seamless information exchange among the various rehabilitation centres participating in Fit4MedRob, adding another layer of complexity and risk. Lastly, the ambitious timeline poses its own set of uncertainties, with doubts about whether the planned activities can lead to a functional prototype within three years, given the project's extensive scope and the relative scarcity of literature on the addressed themes.
Within the Fit4MedRob project framework, the development of the prototype data platform for robot-aided rehabilitation marks a significant stride towards adopting advanced data analysis techniques within healthcare practices. This initiative, driven by the active participation and collaboration of a multidisciplinary consortium of partners, aims to forge a solution that is not only innovative in addressing daily clinical needs but is also deeply grounded in practical application. At the heart of the Fit4MedRob platform project is the ambition to introduce a novel methodology for managing and analyzing data in the rehabilitation sector, aiming to overcome fundamental challenges such as data integration, privacy protection, and advanced data analytics. The initial outcomes of this effort, including the development of a layered architectural framework, the use of open-source technologies, and a focused attention on data interoperability, lay the foundation for ongoing innovation in the realm of healthcare data-driven technologies.

Regarding the project's validation and the assessment of its true effectiveness, the real-world applicability will be demonstrated in the forthcoming pragmatic clinical trials. The trials, commencing in the following months, is set to involve a substantial cohort of rehabilitation patients, recruited nationally, providing a comprehensive evaluation of the platform's impact and effectiveness. This step is crucial for assessing the practical benefits of the platform and its contribution to the rehabilitation ambit, offering an extensive demonstration of how data-driven approaches can enhance patient care and outcomes.

Looking ahead, the healthcare data platform could undergo several enhancements to broaden its capabilities and impact. The potential developments include integrating a wider variety of data from diverse sources such as electronic health records, wearable devices, remote patient monitoring, and smart sensors. This expansion would enable a more comprehensive and detailed view of patients and their conditions, facilitating the identification of more accurate correlations and patterns to improve rehabilitation outcomes.

Incorporating artificial intelligence for diagnosis and treatment planning is another prospective advancement. The platform could integrate AI algorithms and machine learning to assist with diagnostics and crafting personalized treatment plans. Machine learning models, for instance, could be trained using data from the platform to predict patient recovery trajectories, suggest tailored treatment pathways, and provide evidence-based recommendations.

Furthermore, the platform could enhance collaboration and knowledge sharing among various stakeholders in the rehabilitation field, leading to more standardized rehabilitation procedures and improved knowledge exchange among healthcare professionals. As telemedicine and digital health services continue to expand, the platform could adapt to support tele-rehabilitation, allowing patients to receive care and support from the comfort of their homes while integrating data from rehabilitation sessions for continuous monitoring and treatment improvement.

Advanced visualization tools could also be incorporated to clearly and intuitively present aggregated data, recovery trends, and treatment performance, aiding healthcare professionals in quickly assessing outcomes and key success metrics. Lastly, the development of predictive models for the prevention and management of health conditions using data from the platform could enable early identification of patients at risk of developing certain conditions or complications, facilitating timely interventions.

These are just a few of the possible future developments for such a platform. Technological evolution and advancements in rehabilitation research could lead to further innovations and opportunities to enhance rehabilitation outcomes and efficacy.

REFERENCES

[1] TT. Perls, “The different paths to 100”, Am J Clin Nutr, 83(2):484S-487S, February 2006.

[2] MJ. Divo, CH. Martinez, Dm. Mannino, “Ageing and the epidemiology of multimorbidity”, Eur Respir J., 44(4):1055-68, October 2014.

[3] D. Ward, A. Drahota, D. Gal, M. Severs, TP. Dean, “Care home versus hospital and own home environments for rehabilitation of older people,” Cochrane Database Syst Rev., 2008(4):CD003164, October 2008.

[4] SG. Parker, P. Oliver, M. Pennington, J. Bond, et al., “Rehabilitation of older patients: day hospital compared with rehabilitation at home. A randomised controlled trial”, Health Technol Assess., 13(39):1-143, iii-iv, August 2009.

[5] World Health Assembly. Fifty-Eighth World Health Assembly. World Health Organ, 2005.

[6] A. Nyman, M. Zingmark, M. Lilja, S. Guidetti, “Information and communication technology in home-based rehabilitation - a discussion of possibilities and challenges,” Scand J Occup Ther., 30(1):14-20, January 2023.

[7] T. Greenhalgh, J. Wherton, C. Papoutsi, J. Lynch, G. Hughes, et al., “Beyond Adoption: A New Framework for Theorizing and Evaluating Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies”, J Med Internet Res., 19(11): e367, November 2017.

[8] Health Level 7. HL7 Web site. Available at: http://www.hl7.org. Accessed Apr 11, 2024

[9] D. Bender, K. Sartipi, “HL7 FHIR: An agile and RESTful approach to healthcare information exchange”, Proceedings of CBMS 2013 - 26th IEEE International Symposium on Computer-Based Medical Systems, 2013.

[10] S. Sauermann, V. David, A. Schlögl, R. Egelkraut, M. Frohner, et al, “Biosignals, Standards and FHIR - The Way to Go?”, Stud Health Technol Inform., 236:356-362, 2017.

[11] N. Liappas, JG. Teriús-Padrón, RI. García-Betances, MF. Cabrera-Umpiérrez, MT. Arredondo, “User-Centered Design in Defining and Developing Health and Well-Being ICT Solutions”, In:Assisted Living. Computer Communications and Networks. Springer, Cham, pp.191-208, 2020.

[12] C. Délétroz, C. Del Grande, S. Amil, et al., “Development of a patient-reported outcome measure of digital health literacy for chronic patients: results of a French international online Delphi study”, BMC Nurs., 22(1):476, 2023.
[13] JC. Goldsack, CA. Zanetti, “Defining and Developing the Workforce Needed for Success in the Digital Era of Medicine”, Digit Biomark., 4(Suppl 1):136-142, November 2020.
[14] PJ. Jongen, “Information and Communication Technology Medicine: Integrative Specialty for the Future of Medicine”, Interact J Med Res., 12:e42831, July 2023.
[15] E. Coffetti, W. Paans, PF. Roodbol, J. Zuidersma, “Individual and Team Factors Influencing the Adoption of Information and Communication Technology by Nurses: A Systematic Review”, Comput Inform Nurs, 41(4):205–14, April 2022.
[16] A. Janssen, T. Robinson, M. Brunner, et al., “Multidisciplinary teams and ICT: a qualitative study exploring the use of technology and its impact on multidisciplinary team meetings.” BMC Health Serv Res., 18(1):444, June 2018.
[17] P. Wieder, H. Nolte, “Toward data lakes as central building blocks for data management and analysis”, Front Big Data. 19; 5:945720, August 2022.
[18] S. Campagnini et al., "Machine learning methods for functional recovery prediction and prognosis in post-stroke rehabilitation: a systematic review," in Journal of NeuroEngineering and Rehabilitation, vol. 19, no. 1, pp. 54, 2022.
[19] "FIT4MEDROB." [Online]. Available: https://www.fit4medrob.it/. [Accessed: 27-Mar-2024].
[20] "Fit4MedRob - Lider Lab." [Online]. Available: https://www.lider-lab.it/fit4medrob/. [Accessed: 27-Mar-2024].
[21] "FIT4MEDROB: Robotic research and artificial intelligence in the daily use of the Italian healthcare system - Unimore." [Online]. Available: https://www.focus.unimore.it/fit4medrob-la-ricerca-robotica-e-lintelligenza-artificiale-nelluso-quotidiano-del-sistema-sanitario-italiano/. [Accessed: 27-Mar-2024].
[22] M. S. Islam et al., "A Systematic Review on Healthcare Analytics: Application and Theoretical Perspective of Data Mining," in Healthcare (Basel), 2018.
[23] M. Jones et al., "Big Data Analytics and Sensor-Enhanced Activity Management to Improve Effectiveness and Efficiency of Outpatient Medical Rehabilitation," in Int J Environ Res Public Health, vol. 17, no. 3, 748, 2020 Jan 24.
[24] L. Petersson, I. Larsson, J.M. Nygren et al., "Challenges to implementing artificial intelligence in healthcare: a qualitative interview study with healthcare leaders in Sweden," in BMC Health Serv Res, vol. 22, 850, 2022.
[25] M. Äijö, J. Reijula, and P. Komulainen, "eHealth challenges in rehabilitation processes - rehabilitation professionals’ experiences in the North Savo region," in Finnish Journal of EHealth and EWelfare, vol. 14, no. 1, pp. 119–130, 2022.
[26] J. Mitchell, C. Shirota, and K. Clanchy, "Factors that influence the adoption of rehabilitation technologies: a multi-disciplinary qualitative exploration," in J NeuroEngineering Rehabil, vol. 20, 80, 2023.
[27] Y. Rybarczyk, J. Kleine Deters, C. Cointe, and D. Esparza, "Smart Web-Based Platform to Support Physical Rehabilitation," in Sensors, vol. 18, 1344, 2018.
[28] P. Maceira-Elvira, T. Popa, AC. Schmid et al., "Wearable technology in stroke rehabilitation: towards improved diagnosis and treatment of upper-limb motor impairment," in J NeuroEngineering Rehabil, vol. 16, 142, 2019.
[29] "Health Level Seven International. HL7 FHIR Release 4 (Version 4.0.1)," [Online]. Available: https://www.hl7.org/fhir/. [Accessed: 27-Mar-2024].
[30] "National Electrical Manufacturers Association. Digital Imaging and Communications in Medicine (DICOM) Part 1: Introduction and Overview (Standard No. PS3.1-2020)," [Online]. Available: https://www.dicomstandard.org/current/. [Accessed: 27-Mar-2024].
[31] D. Bender and K. Sartipi, "HL7 FHIR: An Agile and RESTful approach to healthcare information exchange," in Proc. 26th IEEE Int. Symp. on Computer-Based Medical Systems, Porto, Portugal, 2013, pp. 326-331.
[32] M. Mustra, K. Delac, and M. Grgic, "Overview of the DICOM standard," in 2008 50th International Symposium ELMAR, vol. 1, 2008.
[33] K. El Emam and F. K. Dankar, "Protecting privacy using k-anonymity," in Journal of the American Medical Informatics Association, vol. 15, no. 5, pp. 627-637, 2013.
[34] N. Mehta and A. Pandit, "Concurrence of big data analytics and healthcare: A systematic review," in International Journal of Medical Informatics, vol. 114, pp. 57-65, 2018.
[35] A. Gonzales, G. Guruswamy, and S. R. Smith, "Synthetic data in health care: a narrative review," in PLOS Digital Health, vol. 2, no. 1, e0000082, 2023.
[36] A. Kebaili, J. Lapuyade-Lahorgue, and S. Ruan, "Deep Learning Approaches for Data Augmentation in Medical Imaging: A Review," in J. Imaging, vol. 9, 81, 2023.

[37] I. Goodfellow et al., "Generative adversarial nets," in Advances in neural information processing systems, vol. 27, 2014.
[38] M. J. Cardoso et al., "MONAI: An open-source framework for deep learning in healthcare," ArXiv abs/2211.02701, 2022.
[39] W. H. L. Pinaya et al., "Generative AI for Medical Imaging: extending the MONAI Framework," arXiv preprint arXiv:2307.15208, 2023.
[40] "Amazon Web Services, Inc. Amazon HealthLake," [Online]. Available: https://aws.amazon.com/healthlake/. [Accessed: 27-Mar-2024].
[41] "Google Cloud. Cloud Healthcare API," [Online]. Available: https://cloud.google.com/healthcare-api. [Accessed: 27-Mar-2024].
[42] "Healthcare on Microsoft Azure," [Online]. Available: https://azure.microsoft.com/en-us/solutions/industries/healthcare/. [Accessed: 27-Mar-2024].
[43] "Oracle Healthcare - Solutions for Healthcare Providers," [Online]. Available: https://www.oracle.com/healthcare/. [Accessed: 27-Mar-2024].
[44] O. Ali, A. Shrestha, J. Soar, and S. Fosso Wamba, "Cloud computing-enabled healthcare opportunities, issues, and applications: A systematic review," in International Journal of Information Management, vol. 43, pp. 146-158, 2018
[45] P.A. Harris, R. Taylor, R. Thielke, J. Payne, N. Gonzalez, and J.G. Conde, "Research electronic data capture (REDCap) – A metadata-driven methodology and workflow process for providing translational research informatics support," in J Biomed Inform, vol. 42, no. 2, pp. 377-381, Apr. 2009.
[46] P.A. Harris, R. Taylor, B.L. Minor, V. Elliott, M. Fernandez, L. O'Neal, L. McLeod, G. Delacqua, F. Delacqua, J. Kirby, S.N. Duda, and REDCap Consortium, "The REDCap consortium: Building an international community of software partners," in J Biomed Inform, May 9, 2019.
[47] "FUME FHIR Converter," [Online]. Available: https://www.fume.health/. [Accessed: 27-Mar-2024].
[48] "HAPI FHIR - The Open Source FHIR API for Java," [Online]. Available: http://hapifhir.io/. [Accessed: 27-Mar-2024].
[49] S. Jodogne, C. Bernard, M. Devillers, E. Lenaerts, and P. Coucke, "Orthanc - A lightweight, restful DICOM server for healthcare and medical research," in Proc. IEEE Int. Symp. on Biomedical Imaging, 2013, pp. 190-193.
[50] R. Hosch, G. Baldini, V. Parmar et al., "FHIR-PYrate: a data science friendly Python package to query FHIR servers," in BMC Health Serv Res, vol. 23, 734, 2023.
[51] J. Walonoski et al., "Synthea: An approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record," in Journal of the American Medical Informatics Association, vol. 25, no. 3, pp. 230–238, Mar. 2018.
[52] Tiangolo, "FastAPI (Version 0.110) [Software]," [Online]. Available: https://github.com/tiangolo/fastapi. [Accessed: 27-Mar-2024].
[53] Celery contributors, "Celery (Version 5.1.2) [Software]," [Online]. Available: https://github.com/celery/celery. [Accessed: 27-Mar-2024]