

OR1 - Architettura della piattaforma e modello dati

Descrizione sintetica dell'obiettivo

Lo scopo dell'OR1, di tipo Ricerca Industriale (RI), è definire una piattaforma innovativa ed all'avanguardia rispetto allo stato dell'arte scientifico e tecnologico, atta a fornire ai cittadini un repository intelligente attraverso sistemi di cloud computing, specializzati per lo storage e l'elaborazione di dati sanitari e clinici. I dati saranno acquisiti da molteplici fonti, come dispositivi wearable (smart-watch e smartphone), sistemi diagnostici e medicali, oltre ad essere inseriti direttamente dall'utente, che potrà sia archiviare digitalmente la propria documentazione sanitaria, sia inserire manualmente testi, in forma di diario o commenti. La piattaforma dovrà essere in grado di acquisire formati di dati che vanno dal testo (strutturato e non), alle immagini biomediche, fino ai segnali biomedici. Oltre alla funzione di data storage, il sistema sarà in grado di fornire servizi a valore aggiunto attraverso l'elaborazione di grosse moli di dati, utilizzando tecniche di Big Data Analytics. In tal modo, sarà possibile ottenere funzionalità di indexing, statistiche e strumenti di monitoraggio e correlazione, orientati sia al singolo individuo, che a gruppi, fornendo strumenti di analisi avanzati anche ad Enti e agenzie operanti sul territorio.

L'OR in esame è stato ottenuto con successo attraverso le attività seguenti.

A1.1 – Analisi delle sorgenti dati

Obiettivo della Attività

L'attività in esame ha come obiettivo l'identificazione dei dati di interesse che la piattaforma dovrà elaborare, analizzandone i rispettivi tipi ed i formati, allo scopo sia di ottenerne i requisiti, sia di individuare i processi e dispositivi di generazione e acquisizione.

Descrizione Attività

L'Attività in esame si è basata sulla analisi dei processi e dei vari ed eterogenei dispositivi disponibili sul mercato al fine di identificare e censire le fonti dati di interesse per il progetto e da poter quindi utilizzare come riferimento durante l'esecuzione delle successive attività progettuali. Durante tale attività sono state quindi identificate le seguenti tipologie di fonti dati:

- dispositivi wearable non sanitari
- dispositivi personali sanitari
- documenti relativi ad attività e prestazioni sanitarie
- diari, note e commenti espressi dagli utenti:
 - diari
 - interviste e schede dedicate
 - social network

riportando, per ciascuna tipologia di fonti dati sopra elencata, esempi concreti di fonti di dati disponibili e diffuse sul mercato. Dato l'intento finale dell'ecosistema BDA4PHR di offrire servizi a valore aggiunti sia al singolo individuo sia ad Enti ed agenzie operanti sul territorio Data, durante l'attività particolare attenzione è stata poi posta nella identificazione e studio delle principali disposizioni normative e gli standard di riferimento per lo scambio, la gestione e, in generale, il trattamento dei dati acquisiti. In tale ambito si giunti quindi alla identificazione puntuale:

- del quadro normativo di riferimento:
 - regolamento europeo GDPR
 - Fascicolo Sanitario Elettronico e Taccuino Personale dell'Assistito
 - Linee guida di Telemedicina
 - etc.
- di protocolli e di standard:
 - per lo scambio dati in ambito medico
 - di comunicazione tra i dispositivi

A conclusione della attività, è stato raggiunto un livello di conoscenza tale da poter infine identificare e definire tutti i processi di acquisizione dati, da sorgenti dati varie ed eterogenee, di interesse per il progetto e da poter quindi utilizzare come riferimento durante l'esecuzione delle successive attività progettuali.

A1.2 – Analisi dei requisiti e definizione del modello funzionale

Obiettivo della Attività

L'attività si concentra sulla analisi dei requisiti e definizione dei modelli funzionali e non funzionali, svolta attraverso un'analisi preliminare dello stato dell'arte. L'analisi dei requisiti funzionali e non funzionali della piattaforma e la definizione del modello funzionale hanno lo scopo di individuare e definire le funzionalità che la piattaforma offrirà, attraverso un'analisi estesa e puntuale dei requisiti di base che la piattaforma deve soddisfare.

Descrizione Attività

L'attività in esame si è basata sulla analisi dei requisiti funzionali e non funzionali della piattaforma ed sulla identificazione dei principali processi di business, di elaborazione dai dati presenti nella piattaforma, da utilizzare come riferimento per la progettazione della piattaforma. Dopo una preliminare analisi sullo stato dell'arte, dei contesti tipici di utilizzo, degli Attori di riferimento, con i loro rispettivi ruoli e responsabilità, l'attività si è basata sulla identificazione e definizione dei processi di riferimento a valle dell'acquisizione dei dati da sorgenti dati varie ed eterogenee. Sono stati quindi definiti in modo formale i processi di riferimento di:

- Analisi dei Big Data
- Richiesta ed Utilizzo Servizi BDA4PHR
- Elaborazione dei dati provenienti da ciascuna delle sorgenti dati identificate nella precedente attività di analisi

A valle di quanto sopra, è stato quindi possibile specificare i requisiti che l'Ecosistema deve poter soddisfare suddivisi per:

- Requisiti Funzionali, indicando cosa l'ecosistema deve fare e come.
- Requisiti Di Interfaccia Utente E Software, specificando come gli utenti finali devono poter interagire con l'ecosistema e come i componenti interni del sistema devono comunicare ed interagire automaticamente tra loro.
- Requisiti Architetture, specificando i componenti ed i pattern architetture da adottare
- Requisiti Di Privacy E Di Sicurezza, affinché il comportamento del sistema sia in linea con il quadro normativo di riferimento, le linee guida di telemedicina, etc.

L'attività è stata infine conclusa identificando la mappatura tra le funzionalità del sistema e lo standard HL7 EHR-S FM.

A1.3 – Definizione del modello dati

Obiettivo della Attività

Lo scopo dell'attività è quello di definire il modello concettuale dei dati usati e gestiti dalla piattaforma, attraverso l'individuazione di elementi, attributi e relazioni dei concetti del dominio di interesse (dati clinici e sanitari individuati nelle precedenti attività), per quello che riguarda la memorizzazione, la gestione, l'indicizzazione e l'elaborazione. Il processo di progettazione del modello concettuale dei dati riguarderà la rappresentazione dei concetti di interesse attraverso un modello formale, ad alto livello, indipendente dagli aspetti implementativi, ma esclusivamente dipendente dalle funzionalità della piattaforma da realizzare.

Descrizione Attività

L'attività in esame si è basata prima su una analisi preliminare e sintetica della tipologia dei dati disponibili da cui derivare ed alimentare il modello dati. Emergendo la presenza di dati sanitari di interesse di natura sia strutturata che non strutturata, è stata quindi identificata come vincente la strategia integrata di approccio ai dati denominata Schema on Read, secondo cui, l'interpretazione dei dati mediante un unico ed integrato modello dati, deve avvenire a valle della loro acquisizione e memorizzazione all'interno della piattaforma. L'attività si è quindi conclusa con l'identificazione del Modello Dei Dati dell'Ecosistema, eseguita attraverso adozione dello standard FHIR (Fast Healthcare Interoperability Resources), scelta ritenuta vincente nella definizione del modello per:

- la disponibilità di risorse FHI semplici, chiare e non ambigue. Le risorse hanno un riscontro concreto e facilmente identificabile nella realtà;
- la possibilità di generare estensioni in modo da garantire copertura anche per i casi limite non ancora contemplati dallo standard;
- l'agevole scambio di informazioni sanitarie attraverso sistemi informativi: le risorse possono essere riusate a prescindere dalla tecnologia di trasporto impiegata;
- la possibilità di completamento degli standard di rilievo in ambito sanitario: FHIR può soddisfare le necessità che erano precedentemente coperte dagli standard di interoperabilità HL7 (V2, V3, CDA);
- l'approccio standard adatto anche in caso di applicazioni di natura mobile.

A1.4 – Individuazione dei moduli della piattaforma

Obiettivo della Attività

La piattaforma BDAPHR dovrà essere modulare e scalabile e supportare gli strumenti di indexing e di analisi avanzati previsti. In tale contesto, l'attività in esame intende individuare e definire i singoli componenti della piattaforma, i quali devono offrire, tra le altre, le funzionalità di:

- repository su cloud
- indicizzazione e reperimento informazioni
- Big Data Analytics
- elaborazione di segnali e immagini biomediche
- comunicazione e trasmissione sicura
- visualizzazione output
- interfaccia utente

Descrizione Attività

L'attività in esame ha consentito di introdurre una visione white/grey box della piattaforma BDA4PHR. Sono stati infatti identificati:

- per la realizzazione del repository su cloud:
 - componenti NoSQL, relazionali e componenti di gestione e monitoraggio del repository;
- per l'indicizzazione ed il reperimento informazioni:
 - componenti di interfaccia, di mapping, di estrazione, di raggruppamento e di costruzione;
- per Big Data Analytics:
 - componenti per data integration ed aggregation e componenti di estrazione di statistiche avanzate - forecasting, pattern recognition, classificazione, etc.;
- per l'elaborazione di segnali e immagini biomediche:
 - componenti per il calcolo del rapporto segnale rumore di un'immagine, di segmentazione delle strutture di interesse in un'immagine, di estrazione di features del primo ordine, di filtraggio di segnali biomedici, di applicazione di tecniche di machine learning e deep learning;
- componenti di comunicazione e trasmissione sicura
- per la realizzazione delle funzionalità di visualizzazione dell'output:
 - componenti di esplorazione, di discovery, di report e di delivery/condivisione;
- per la realizzazione della interfaccia utente:
 - componenti GUI per applicazioni web e per applicazioni mobile.

Per ogni componente identificato, sono state poi identificati i corrispettivi modelli e le tecnologie di riferimento.