

TITOLO
SECSHARE-NGS: Piattaforma per la condivisione dei dati clinici e molecolari di Next Generation Sequencing nel rispetto della sicurezza e privacy

Soggetto/i Attuatore/i (Soggetto/i Partner coinvolti nel progetto)

- **UNLIMITED TECHNOLOGY SRL**
- **BIOGEM (Biologia e Genetica Molecolare) consorzioM Data System s.r.l.**

Il progetto si colloca nell'ambito dello sviluppo di **nuove tecnologie ICT a supporto della diagnostica molecolare e medicina personalizzata**. In particolare, la disponibilità di sistemi di indagine basati su informazioni molecolari (*Next Generation Sequencing*) pone numerose sfide soprattutto di carattere computazionale, e richiede l'adozione di nuovi modelli di protezione, archiviazione e condivisione dei dati. Scopo del progetto è quello di mettere a punto un sistema di condivisione dei dati genomici che possa affrontare alcuni dei limiti attuali quali:

- **necessità di scalabilità** compatibile con le sfide dei *big data*: le attuali tecnologie di *Next Generation Sequencing* (NGS) ci hanno ormai proiettato nell'era della *personalized genomics* in cui tutti potranno avvalersi delle proprie informazioni molecolari per definire i protocolli di cura e predire eventualmente la risposta e l'efficacia di farmaci in base al proprio profilo molecolare. Tuttavia le tecnologie NGS producono quantità notevoli di dati (diversi Tbytes per ogni singolo esperimento) che devono essere rappresentati, condivisi ed interrogati per usi secondari e studi di associazione.
- **Rispetto della sicurezza e privacy** nella gestione dei dati genomici è necessario tenere in considerazione che, oltre alle problematiche di sicurezza e privacy già presenti nell'ambito della gestione dei dati clinici, non è possibile la semplice anonimizzazione attraverso la rimozione dei dati identificativi dai sensibili. Molti studi recenti hanno infatti dimostrato che dagli stessi dati genomici è possibile risalire all'identità del paziente, mettendo in evidenza che l'identità è già contenuta nel dato molecolare. Pertanto è necessario mettere in campo nuovi modelli disegnati *ad hoc* per la identificazione e la protezione di dati genomici. Un ulteriore aspetto da tenere presente, nell'ambito dei sistemi NGS è che essi forniscono uno spettro di informazioni su larga scala (ad esempio l'*exome sequencing* fornisce tutte le varianti individuali sulle zone codificanti del genome, ed addirittura il *whole genome sequencing* riporta l'intero genoma) che devono poter essere condivise solo in parte. Quindi deve essere il paziente a definire quali parti del genoma, in base alla patologia, possono essere visionate da operatori sanitari, mentre potrebbe volerne mascherare altre (ad esempio varianti che predispongono a specifiche patologie (come il gene ApoE per l'Alzheimer o i geni BRCA per il cancro al seno e molti altri). L'aspetto temporale è particolarmente importante anche perché queste tipologie di informazione, a differenza dei classici dati clinici, non cambiano durante la vita del paziente, e paradossalmente coinvolgono le scelte e l'identità dei familiari del paziente. Alcuni casi clamorosi in America hanno infatti evidenziato come la divulgazione dei dati di un particolare paziente usato per numerosi studi scientifici legato al cancro del collo dell'utero hanno poi coinvolto i discendenti della paziente in dispute legali (Skloot, R. (2013). *The immortal life of Henrietta Lacks, the sequel*.)
- **Facilità di integrazione con i sistemi di gestione dei dati clinici nel rispetto degli standard internazionali**. La standardizzazione dei dati clinici è uno dei temi fondamentali per garantire e semplificare la continuità delle cure, per consentire il miglioramento della spesa sanitaria e talvolta anche per migliorare l'efficacia e l'appropriatezza delle cure. L'adozione di standard internazionali (quali ad es. HL7/CDA) è considerato un requisito vincolante per qualsiasi sistema di gestione dei dati clinici. Esistono attualmente diversi consorzi internazionali che hanno l'analogo l'obiettivo di standardizzare la rappresentazione dei dati molecolari. Quello forse maggiormente avanzato è sicuramente il consorzio eMERGE

(Electronic Medical Records and Genomics Network).

Il progetto prevede la progettazione e realizzazione di un sistema di gestione e condivisione dei dati NGS tenendo conto degli aspetti menzionati in precedenza. Le competenze e le expertise del gruppo di lavoro sono decisamente adeguate al raggiungimento di questo scopo e fanno anche tesoro di attività di collaborazione che da diversi anni sono portate avanti individualmente ed in collaborazione fra i vari partner. In particolare la Unlimited Technology s.r.l. è parte del gruppo Unlimited Software che ha sviluppato numerose esperienze in questo campo, ad esempio il progetto KON-CUBE (<http://www.koncube.org>) finanziato dalla Regione Campania misura 3.17 ha visto lo sviluppo di un sistema di supporto alle decisioni nell'ambito della medicina personalizzata in oncologia, ulteriori attività sono svolte nell'ambito del progetto CAMPUS-BIOFRAME (in collaborazione con il centro di ricerca CEINGE) che ha per obiettivo la Realizzazione di una piattaforma integrata per studi genomici e trascrittomici di processi di riparo tissutale in cui l'attenzione è focalizzata sugli aspetti funzionali e di analisi dei dati genomici. Entrambe le aziende sono inoltre presenti sul mercato da diversi anni nella realizzazione di sistemi di gestione dei dati sanitari e nella consulenza nell'ambito ICT per la Sanità.

Sempre in collaborazione con Unlimited Software è stata realizzata in modo completo una Certification Authority, questa CA è l'unica certificata in ambito pubblico per lo stato egiziano.

Lo studio delle informazioni genomiche è inoltre la missione principale del centro di ricerca BIOGEM che al suo interno vede la presenza di un importante laboratorio di Bioinformatica e di un Laboratorio di Genomica. Nell'ambito delle attività di questi laboratori si svolgono ricerche testimoniate da numerose pubblicazioni in ambito internazionale e di diversi prototipi che hanno come oggetto lo sviluppo e l'applicazione di nuove tecnologie software per l'analisi di dati genomici. In particolare, per il progetto SEC SHARE-NGS, il laboratorio ha messo a punto diversi prototipi per l'analisi di dati NGS in ambito oncologico legati all'associazione di mutazioni ed alterazioni genomiche a sottotipi di tumori ed individuazione di specifici marcatori prognostici. Gli studi in tali ambito hanno evidenziato alcuni aspetti significativi legati alla gestione e condivisione di questo specifico tipo di dati e sono state progettate alcune soluzioni ad hoc basate su crittografia omomorfa e sistemi di crittografici a chiave pubblica.

Il progetto prevede infatti il trasferimento delle competenze e metodologie di rappresentazione analisi di dati genomici sviluppate nel Laboratorio di Bioinformatica di BIOGEM in ambito industriale per la realizzazione del sistema di condivisione che sarà descritto nei paragrafi successivi.

La soluzione tecnologica individuata permette di effettuare un salto tecnologico e quindi mette in condizione l'azienda proponente di diventare un player di riferimento all'interno di questa tecnologia emergente. Il sicuro sviluppo e diffusione delle informazioni genetiche personali è anticipato da vari fattori ed esigenze oramai chiare ed evidenti. La piattaforma di sviluppo proposto sarà all'avanguardia e permetterà sia alla soluzione che all'azienda di avere

una visibilità a livello mondiale. I possibili utenti possono essere ricercati a più livelli in quanto la sua proposizione giungerà sia al singolo utente che agli ospedali, enti ricerca e enti governativi.