

PIANO DI LAVORO

Soggetti attuatori

Soggetto proponente

Denominazione e ragione sociale CRS4			
Sede Legale			
Loc. Piscinamanna	Cap. 09010	Città Pula	Prov. CA
Sede Operativa			
Loc Piscinamanna	Cap. 09010	Città Pula	Prov. CA
Nome e Cognome del legale rappresentante PAOLO ZANELLA			

Introduzione: Progetto di Bioinformatica

All'interno del distretto tecnologico della biomedicina con particolare concentrazione nell'area Cagliari – Pula e nel Parco Scientifico e Tecnologico della Sardegna sono insediate imprese e centri di ricerca che si occupano della produzione, con esperimenti di laboratorio, di dati biologici, genetici e farmacologici.

Il lavoro in laboratorio è seguito da un'attività di analisi dei dati al computer, con l'uso di strumenti bioinformatici, come database e software dedicati; alcuni strumenti sono accessibili gratuitamente dal web, alcuni sono prodotti commerciali.

Centri specializzati in bioinformatica, attraverso siti web, rendono disponibili dati e strumenti di analisi, a cui possono accedere tutti i ricercatori in possesso di una connessione Internet. I centri più importanti si trovano negli Stati Uniti (NCBI), in Gran Bretagna (EBI) e in Giappone (DDBJ).

Da qualche mese le imprese biomediche e i centri di ricerca del territorio possono accedere agli strumenti bioinformatici attraverso un BioPortale realizzato all'interno del Parco, dal CRS4, nell'ambito del Progetto di Bioinformatica e medicina personalizzata.

Il gruppo di Bioinformatica del CRS4 è nato nel 2005 e il BioPortale è stato messo in rete nella primavera del 2006: si trova in una fase iniziale di costruzione e implementazione, pur essendo già attivo.

Pur non potendo raggiungere i livelli dei siti realizzati dai grandi gruppi di importanza internazionale, il BioPortale del CRS4 può fare la differenza per le organizzazioni del territorio, grazie alla vicinanza fisica di produttori e fruitori dei servizi.

La possibilità di poter strettamente interagire con gli utenti del portale consente al gruppo del CRS4 di offrire delle soluzioni su misura, progettate in base alle esigenze dei ricercatori e dei loro studi.

Lo scopo di questo progetto cluster è quello di realizzare strumenti bioinformatici, accessibili dal BioPortale, che rispondano perfettamente alle richieste delle imprese e degli altri gruppi di ricerca locali. Raggiungere un alto livello nell'offerta dei servizi può contribuire ad accrescere la competitività del territorio e più specificamente degli operatori del distretto biomedico.

Breve descrizione del Laboratorio di Bioinformatica del CRS4.

Il gruppo di Bioinformatica del CRS4 è impegnato in due attività:

- ricerca
- servizi&sviluppo.

La ricerca si svolge su tre principali aree di studio: la genomica, le biotecnologie e la biologia dei sistemi. Il personale impegnato nella ricerca mette a disposizione il suo background di conoscenze ed esperienza anche per la produzione dei servizi.

L'attività di produzione di servizi&sviluppo prevede principalmente la costruzione, lo sviluppo e il continuo aggiornamento di un BioPortale: un sito web che permette l'accesso dall'esterno ai software e ai database di bioinformatica. Il portale è stato progettato per essere utilizzato con facilità da ricercatori con formazione biologico-medica, senza esperienza in informatica.

Tecnicamente, quando un utente seleziona un'applicazione e sceglie le opzioni per avviarla, il software viene eseguito su un cluster da 44 nodi del CRS4, chiamato Janas. Il cluster fornisce la necessaria capacità di elaborazione per programmi con alto costo computazionale.

Il lavoro sul portale prevede:

- la realizzazione di una copia in locale delle principali banche di dati e software di biologia molecolare con la tecnologia di mirroring. Giornalmente sono "mirrorizzate" le banche dati da quattro differenti siti (NCBI in USA, EBI and Ensembl in UK, ExPasy in Svizzera) e sono rese disponibili in locale dal server FTP anonimo al <ftp.bioinformatica.crs4.org>.
- l'installazione di software bioinformatici sui pc del CRS4 e le eventuali modifiche. I software attualmente installati generalmente includono:
 - software di analisi di sequenza (DNA, RNA e proteine)
 - software di modellizzazione (proteine)
 - accesso ai database (GO, Ensemble).

Progetto cluster di bioinformatica

Obiettivi

In linea con i progetti cluster del Consorzio Ventuno anche questo progetto ha l'obiettivo di consentire alle imprese, università e centri di ricerca che vi aderiscono di arrivare a un rafforzamento tecnologico attraverso l'adozione di soluzioni innovative nel campo delle tecnologie e dei tools bioinformatici.

Il principale obiettivo del progetto cluster è infatti quello di arricchire il portafoglio servizi offerti, attraverso il BioPortale del CRS4, con strumenti bioinformatici scelti in base alle esigenze degli utenti, per aree di interesse molto innovative, ancora in via di definizione.

Con il progetto si mira a diffondere tra gli operatori del comparto processi di innovazione e di sviluppo che possono scaturire dalla condivisione e dall'approfondimento delle conoscenze di settore, sul piano tecnico-specialistico. In quest'ottica il progetto rappresenta un'occasione di networking a livello locale per facilitare lo scambio di conoscenza e la creazione di una comunità di attori in grado di sviluppare progetti e tecnologie innovative nei campi della bioinformatica.

Seminari e workshop formativi saranno usati come occasioni d'incontro e di scambio di informazioni tra gestori del sito e utenti: da una parte i ricercatori, in maggioranza di formazione biologica, potranno formarsi sul funzionamento dei nuovi strumenti installati sul sito, dall'altra i bioinformatici conosceranno da vicino le esigenze dei biologi per proporre le soluzioni più adatte.

Struttura

Il progetto prevede principalmente lo sviluppo di applicazioni bioinformatiche, che avverrà per cinque differenti aree di interesse, anche queste selezionate considerando le tematiche di ricerca in cui sono attualmente impegnate le organizzazioni e le imprese del territorio e quelle che saranno le tematiche del prossimo futuro:

1. Database
2. Strumenti per gestione e analisi di dati genetici e genealogici
3. Strumenti per gestione e analisi di dati di MicroArray
4. Strumenti per gestione e analisi di dati di proteomica
5. Data and text mining.

La caratteristica principale delle applicazioni bioinformatiche sviluppate è di essere adattate ai bisogni dei ricercatori delle imprese biomediche locali.

Per raggiungere questo obiettivo il progetto prevede per ogni area tematica:

- incontri con la comunità utenti
- workshop e corsi di formazione su strumenti e tecnologie
- la collaborazione tra gestori del BioPortale e gli utenti per la progettazione e lo sviluppo dei software bioinformatici
- la produzione di documentazione scritta per i software studiati e installati durante la realizzazione del progetto.

Anche lo svolgimento dell'attività di formazione prevede un continuo adattamento alle richieste degli utenti. A questo scopo sono previsti incontri di due ore ciascuno ogni due mesi per discutere e decidere sugli argomenti da trattare, scegliere gli esperti da invitare per i workshop e valutare eventuali proposte dei ricercatori.

Descrizione del progetto

I primi cinque pacchetti di lavoro (work package, WP), che verranno di seguito descritti, sono stati pensati in base a una divisione per area di interesse (vedi struttura del progetto). Un sesto pacchetto di lavoro è invece relativo alla gestione del progetto e alle infrastrutture.

WP1: Database

I database rappresentano uno strumento indispensabile per i ricercatori in biologia molecolare, per l'archiviazione e l'organizzazione della grande quantità di dati che viene prodotta giornalmente nei laboratori di tutto il mondo.

Il primo passo previsto nel progetto per quest'area di interesse è l'organizzazione di un workshop per informare gli utenti del BioPortale sull'argomento. Grazie all'incontro gli utenti potranno essere in grado di operare una scelta ragionata sul sistema da adottare per il BioPortale tra quelli esistenti.

Argomenti trattati nel workshop:

1. introduzione alle diverse teorie e tecnologie dei database;
2. teoria del database relazionale;
3. tecnologia del object oriented database;
4. progettare database: UML e Entity relationship;
5. introduzione al linguaggio SQL;
6. strumenti di programmazione per l'accesso ai database (Perl, java, Python, PHP);
7. presentazione dei sistemi di database più utilizzati in biologia molecolare: Oracle, MySQL, PostgreSQL.

Questa fase del progetto è essenziale per la realizzazione dei successivi pacchetti di lavoro e deve essere quindi realizzata per prima. I concetti che saranno trasmessi su questo argomento, infatti, sono la base indispensabile per i workshop, i corsi di formazione e le attività di sviluppo delle altre aree di interesse.

Durata

Preparazione: 1 mese

Workshop: 3 giorni

Risultati attesi

Gli utenti potranno acquisire una comprensione delle differenti tecnologie di database usate in biologia molecolare e quindi scegliere quale tecnologia possa essere più adatta alle proprie esigenze.

Deliverables

Workshop sulle tecnologie dei database con relativa documentazione scritta.

WP2: Analisi dei dati genetici e genealogici

I progetti di analisi genetica di casi familiari richiedono la gestione di informazioni genealogiche e cliniche. La soluzione ideale è salvare le informazioni all'interno di un database, dotato di accesso controllato alle informazioni riservate sui pazienti. Data la natura riservata delle informazioni, esiste la necessità di sviluppare un tipo di database che ciascun gruppo possa installare e mantenere in locale. Questo sistema dovrà poi essere collegato a un programma per il disegno dei pedigree e a un software di analisi dei dati genealogici. E' anche da valutare la possibilità di mantenere un repository centralizzato con dati pubblici e una cartella condivisa che riporti quali dati sono disponibili e accessibili in ciascun centro e il nome della persona da contattare per l'accesso.

Personale:

WP2.1 e WP2.2: 1 esperto di database per un impegno full time, che dovrà anche supervisionare il lavoro del programmatore per i punti WP2.3 e WP2.4.

WP2.3 e WP2.4: 1 programmatore (java, C)

WP2.1: Valutazione dei database per i dati genetici e genealogici

E' necessaria la raccolta di informazioni sui database esistenti per la gestione dei dati genetici e genealogici e la stesura del relativo rapporto. I database di questo tipo devono poter contenere la storia familiare, i dati di genotyping e analisi di linkage.

Un workshop sarà organizzato per presentare i database che rispondono a queste caratteristiche e illustrare le differenze tra le diverse opzioni. Il workshop sarà anche un'occasione, per i gestori del BioPortale, per interagire con gli utenti e capire le loro particolari esigenze. Alla fine del workshop, utenti e bioinformatici definiranno insieme le modifiche necessarie per implementare il database valutato come il più adatto per i loro lavori di ricerca.

Durata

Valutazione e rapporto: 15 giorni

Preparazione del workshop: 15 giorni

Workshop: 1 giorno.

Risultati attesi

Alla conclusione di questa fase gli utenti sapranno decidere quali modifiche dovranno essere apportate al database scelto fra quelli a disposizione.

Deliverables

Un rapporto sullo stato dell'arte dei sistemi di database per questo tipo di dati. Un workshop per uno scambio di informazioni tra gestori del BioPortale e gli utenti. Uno schema temporale delle modifiche che dovranno essere apportate, dove sarà specificata l'importanza di ciascuno step.

WP2.2: Sviluppo dei database per i dati genetici e genealogici

In seguito alla valutazione e al workshop descritti al punto WP2.1, sarà sviluppato un database che si interfacerà con il software scelto dagli utenti come descritto ai punti WP2.3 e WP2.4.

Se gli utenti sceglieranno di fornire dati pubblici a un repository centrale, sarà sviluppato e implementato anche questo sistema.

I database dovranno essere in grado di:

1. Importare/esportare dati dai programmi di pedigree: gli utenti dovranno fornire la lista dei software più utilizzati che il database dovrà riconoscere.
2. Importare/esportare dati da/a i programmi di QTL, linkage e genotyping: gli utenti forniranno la lista di software da connettere al database.

Durata

Progettazione e sviluppo dei database: 2 mesi

Implementazione dei database e formazione degli utenti: 8 mesi

Risultati attesi

Gli utenti avranno a disposizione i database adatti a gestire i dati prodotti con gli studi familiari e con le analisi di genotyping e linkage.

Deliverables

Un sistema di database per dati genealogici installato nei laboratori degli utenti. Se gli utenti sceglieranno di fornire dati pubblici a un repository centrale, sarà sviluppato e implementato anche questo sistema.

WP2.3: Valutazione del software per la costruzione di pedigree

I software per la costruzione di pedigree sono usati per disegnare l'albero genealogico delle famiglie e mostrare graficamente i risultati degli studi di genotyping e le informazioni cliniche. I software per pedigree più usati comunemente sono "Cyrillic" (solo su PC) e "PedDraw" (solo su Macintosh), tutti e due prodotti commerciali che possono essere modificati solo rivolgendosi direttamente al produttore. Il primo passo sarà dunque la valutazione di tutti i software per pedigree disponibili per cui sia accessibile il codice sorgente per poter apportare delle modifiche.

Un workshop sarà organizzato per presentare i software che rispondono a queste caratteristiche, illustrando le differenze tra le diverse opzioni. Il workshop sarà anche un'occasione, per i gestori del BioPortale, per interagire con gli utenti e capire le loro particolari esigenze. Alla fine del workshop, utenti e bioinformatici definiranno insieme le modifiche necessarie per implementare il software valutato come il più adatto per i loro lavori di ricerca, che sarà poi connesso al database descritto al punto WP2.2.

Durata

Valutazione e rapporto: 14 giorni

Preparazione del workshop: 14 giorni

Workshop: 1 giorno

Risultati attesi

Gli utenti avranno una visione complessiva di tutti i software per pedigree al momento disponibili e avranno gli strumenti necessari per scegliere il software più adatto.

Deliverables

Un rapporto scritto e un workshop sui software per pedigree già esistenti. La scelta degli utenti del software per pedigree da installare e adattare.

WP2.4: Sviluppo dei software per pedigree

Dopo la valutazione e il seminario, il software scelto sarà sviluppato e adattato:

1. il software dovrà essere connesso al database sviluppato al punto WP2.2. Gli utenti dovranno essere in grado di inserire i dati familiari nel database usando il software per pedigree e di visualizzare i dati dal database e di esportarli come immagini;
2. gli utenti dovranno poter visualizzare graficamente l'output del loro programma di QTL, linkage o genotyping. Dovranno inoltre poter esportare i dati delle famiglie nel formato corretto per poterli caricare in questi programmi.

Gli utenti forniranno una lista dei software scientifici di cui hanno bisogno per connettersi al programma per pedigree.

Durata

Sviluppo di una prima versione: 3 mesi

Sviluppo della versione definitiva, con le connessioni ai software come definiti dagli utenti: 8 mesi

Risultati attesi

Un programma per disegnare pedigree corrispondente alle esigenze degli utenti e interconnesso al database descritto al punto WP2.2. Il programma per pedigree dovrà anche importare/esportare dati da una lista di programmi definita dagli utenti.

Deliverables

Un software per il disegno di pedigree per dati genetici, interconnesso con il database descritto al punto WP2.2 e altri software definiti dagli utenti.

WP3: MicroArray

La tecnologia MicroArray è sempre più utilizzata per la ricerca sull'espressione genica e il genotyping. Con l'uso di questa tecnologia è possibile produrre un'enorme mole di dati in poco tempo, dati che devono essere archiviati e analizzati.

Personale: WP3.1, WP3.2: 1 bioinformatico esperto su database dal gruppo del CRS4. Un esperto di database farà da supervisore.

WP3.3, WP3.4: 1 programmatore. L'amministratore del BioPortale farà da supervisore.

WP3.1: Valutazione dei sistemi di database esistenti usati per i dati di MicroArray

Sarà inizialmente valutato lo stato dell'arte dei sistemi usati per archiviare i dati di MicroArray.

Alcuni sistemi:

- Il database ArrayExpress dell'EBI, che raccoglie tutti i dati pubblici e consente l'accesso ai ricercatori;
- il sistema Dbase il cui uso si sta sempre più diffondendo in Europa;
- il Stanford MicroArray Database (SMD) di Stanford negli USA.

I risultati di questo processo di valutazione saranno presentati durante due giorni di workshop, a cui saranno invitati a partecipare alcuni esperti del settore. Il gruppo di lavoro interagirà con gli utenti interessati per capire quali sono le loro esigenze circa l'installazione in locale di un sistema.

Durata

Valutazione e rapporto: 1 mese

Preparazione del workshop: 2 settimane

Workshop: 2 giorni

Risultati attesi

Agli utenti saranno presentati tutti i database per dati di MicroArray disponibili perchè possano decidere quale sistema si adatta meglio alle loro esigenze, considerando il tipo di macchine da laboratorio presenti nel Parco e il tipo di dati sperimentali che producono.

Deliverables

Un rapporto scritto e un workshop sui database per i dati di MicroArray. Alla fine del workshop gli utenti avranno una visione chiara di ciò che dovrebbe essere implementato.

WP3.2: Installazione e adattamento del database per dati di MicroArray

Il database scelto dagli utenti sarà installato e adattato. L'adattamento include:

- import/export dei dati dal/al sistema Affymetrix situato nei laboratori di genotyping
- import/export dei dati dal/al sistema per l'espressione genica situati nei laboratori degli utenti
- import/export dei dati da/ai differenti pacchetti di software usati dagli utenti, che forniranno una lista dei software di cui hanno bisogno per connettersi al database.
- Connessione di R e dei pacchetti di Bioconductor al database.

Durata

Creazione del modello e suo sviluppo: 3 mesi

Implementazione, connessione di tutti i software e formazione: 9 mesi

Risultati attesi

Un database per MicroArray utile alle esigenze degli utenti e connesso ai software indicati dagli utenti stessi.

Deliverables

Un database per l'archiviazione dei dati di MicroArray.

WP3.3: Valutazione sull'analisi dei dati di MicroArray

Sarà effettuata una valutazione su tutti gli strumenti disponibili per analizzare i dati di MicroArray e sarà prodotto un rapporto sullo studio. Durante un workshop saranno presentati i risultati della valutazione e gli utenti saranno formati da esperti del settore sul funzionamento di alcuni software.

Tra gli strumenti più usati:

1. R e bioconductor packages

2. SimWalk
3. Dchip

Durata

Valutazione e rapporto: 1 mese

Preparazione del workshop: 3 settimane

Workshop: 2 giorni

Risultati attesi

Gli utenti definiranno e decideranno quale software è il più adatto alle loro esigenze. Questo sarà installato come descritto al punto WP3.4.

Deliverables

Un rapporto scritto e un workshop sugli strumenti informatici esistenti.

WP3.4: Installazione dei software di analisi dei dati di MicroArray

Il software scelto dagli utenti sarà installato e sarà valutato attraverso l'interazione con gli stessi utenti anche per capire quali tipi di modifiche sono necessarie.

Durata

Installazione di tutti gli strumenti: 1 mese

Sviluppo dell'interfaccia web: 3 mesi

Formazione e adattamenti specifici: 3 mesi

Risultati attesi

Gli strumenti di analisi dei dati di MicroArray scelti dagli utenti saranno installati sui computer di bioinformatica e saranno accessibili dal BioPortale, per sfruttare la velocità di elaborazione delle macchine del CRS4.

Deliverables

Un set di strumenti di analisi di dati di MicroArray, installati sul BioPortale, e la documentazione necessaria per il loro utilizzo.

WP4: Proteomica

Il sequenziamento completo del genoma umano non ha fornito la spiegazione per tutti i dubbi sul suo funzionamento, anzi, ha contribuito al sorgere di nuove domande. Il genoma umano contiene circa 25 000 geni e probabilmente le informazioni per la sintesi di più di un milione di proteine. La proteomica è lo studio del proteoma, ossia il complesso di proteine espresse da un genoma. Le proteine possono essere analizzate usando strumenti per l'analisi di sequenza e tecnologie per la modellizzazione molecolare, che sono già presenti nel BioPortale. Altri metodi di studio hanno come scopo l'identificazione delle proteine in tessuti specifici, e utilizzano le tecnologie della spettrometria di massa.

Questo progetto prevede la continua interazione tra gestori del BioPortale e gruppi di ricerca interessati all'implementazione di database e di software specifici per gli studi di proteomica. La proteomica è una scienza emergente e molti degli strumenti sono ancora in fase di sviluppo, quindi questa parte del progetto sarà continuamente aggiornata per stare al passo con le prossime scoperte e innovazioni tecnologiche.

Personale:

WP4.1, WP4.2: 1 programmatore di database. 1 esperto di database per la supervisione.

WP4.3, WP4.4: 1 un bioinformatico dal gruppo del CRS4 per l'installazione del software. L'amministratore del BioPortale farà da supervisore per la connessione al BioPortale.

1 programmatore per il software di analisi di immagini (se gli utenti dovessero decidere di adottare tale soluzione). Il Dr. Alan Scheinine del CRS4 farà da supervisore.

WP4.1: Valutazione del database per i dati di proteomica

Sarà condotta una valutazione sullo stato dell'arte dei database per la proteomica. Alla fine verrà steso un rapporto che sarà poi presentato in occasione di un workshop. L'interazione con gli utenti sarà fondamentale per pianificare lo sviluppo e l'implementazione del database e degli strumenti correlati.

Durata

Valutazione e stesura del rapporto: 2 settimane

Preparazione del workshop: 1 settimana

Workshop: metà giornata

Risultati attesi

La conoscenza dello stato dell'arte dei database usati per i dati di proteomica. Questo aiuterà gli utenti nella decisione e nella definizione delle proprie esigenze. Gli utenti dovranno comunicare cosa deve essere implementato al punto WP4.2.

Deliverables

Un rapporto scritto e un workshop sui database esistenti usati per la proteomica.

WP4.2: Installazione del database per i dati di proteomica

Il lavoro comincerà con lo sviluppo del database e con l'implementazione degli strumenti definiti dagli utenti.

Durata

Progettazione del modello e sviluppo: 3 mesi

Implementazione e formazione: 8 mesi

Risultati attesi

Un database per i dati di proteomica come definito dagli utenti secondo quanto detto al punto WP4.1.

Deliverables

Un database per i dati di proteomica.

WP4.3: Valutazione dei software per l'analisi dei dati di proteomica

Sarà condotta una valutazione di tutti i software esistenti per l'analisi dei dati di proteomica e i risultati della valutazione saranno riportati su documentazione. Sarà organizzato un workshop per la presentazione degli strumenti, al seguito del quale gli utenti dovranno effettuare una scelta: se utilizzare un software già esistente così com'è o se adattarlo per connetterlo al database per i dati di proteomica di cui si è parlato al punto WP4.2.

Un particolare tipo di software utilizzato negli studi di proteomica è quello per l'analisi di immagini. Nessuno dei software per analisi d'immagine esistenti risponde completamente alle esigenze dei ricercatori. Agli utenti verranno descritti i diversi pacchetti disponibili e le possibili implementazioni, in modo che possano decidere su quale software far apportare le modifiche.

Durata

Valutazione e rapporto: 3 settimane

Preparazione del workshop: 1 settimana

Workshop: metà giornata

Risultati attesi

Agli utenti sarà presentata la situazione dei software attualmente disponibili per l'analisi dei dati di proteomica, perché possano scegliere quale strumento debba essere implementato sul BioPortale.

Deliverables

Sarà preparato un rapporto scritto e si terrà un workshop sugli strumenti bioinformatici per la proteomica.

WP4.4: Installazione del software per l'analisi dei dati di proteomica

Saranno installati e adattati gli strumenti di analisi di dati di proteomica scelti dagli utenti.

Durata

Installazione di tutti gli strumenti: 1 mese

Sviluppo dell'interfaccia web: 3 mesi

Formazione e adattamenti specifici: 3 mesi

Risultati attesi

Gli strumenti scelti dagli utenti saranno installati sui computer del CRS4 per sfruttare la loro capacità di elaborazione.

Deliverables

Un set di strumenti per l'analisi di dati di proteomica e la documentazione necessaria per il loro corretto utilizzo saranno installati sul BioPortale.

WP5: Data e text mining

L'attività di ricerca in biologia molecolare produce giornalmente una grande quantità di informazioni, per cui diventa sempre più difficile per i ricercatori tenersi aggiornati sugli sviluppi più recenti relativi alla loro area di studio. L'informatica sta sviluppando la tecnologia per cercare velocemente e semplicemente grandi quantità di informazioni sia tra le pubblicazioni scientifiche, sia nelle banche dati o nei siti web (text e data mining).

WP5.1: Valutazione su data e text mining

E' prevista una valutazione sulle tecnologie esistenti per il text e data mining. Seguirà la stesura di un rapporto e l'organizzazione di un workshop a cui saranno invitati alcuni esperti del settore.

Durata

Valutazione e rapporto: 1 mese e mezzo

Preparazione del workshop: 2 settimane

Workshop: 1 giorno

Risultati attesi

Gli utenti saranno informati sulle tecnologie per il text and data mining.

Deliverables

Un rapporto scritto sulle tecnologie per il text and data mining e un workshop.

Personale:

1 persona dal gruppo di bioinformatica del CRS4 sarà coinvolta in questo progetto

WP6: Infrastrutture e project management

WP6.1: Infrastrutture

Le infrastrutture necessarie per la realizzazione del progetto si integreranno con quelle già esistenti per il lavoro del gruppo di Bioinformatica del CRS4.

Nel dettaglio serviranno specifiche infrastrutture per

- i database:

- database development and production servers;
- database software.

- lo sviluppo degli strumenti bioinformatici:

- un development server

- il BioPortale:

- un nuovo server (attualmente il BioPortale si trova in un piccolo computer che non sarebbe in grado di supportare gli strumenti per i dati di MicroArray e di proteomica)

- la programmazione:

- un personal computer per ogni programmatore impegnato nel progetto.

WP6.2: Management

Il progetto sarà diretto dalla dottoressa Patricia Rodriguez-Tomé, direttrice del gruppo servizi di Bioinformatica del CRS4. La dottoressa Rodriguez-Tomé ha 19 anni di esperienza nel campo della bioinformatica, e nella direzione e collaborazione in progetti europei e internazionali, oltre che nell'insegnamento in corsi di formazione.

La dottoressa Giuliana Brunetti prenderà parte all'organizzazione dei workshop e dei corsi formativi: dopo la laurea in biologia la dottoressa Brunetti ha frequentato un master post-universitario in divulgazione scientifica e ha fatto esperienza nell'insegnamento di materie scientifiche.

Un programmatore sarà assunto per adeguare e sviluppare continuamente il software per gli studi di pedigree.

Un programmatore sarà assunto per partecipare all'installazione di diversi strumenti e alle loro eventuali modifiche. La supervisione sarà affidata a un membro del gruppo di Bioinformatica che parteciperà al lavoro di installazione.

Un programmatore per database sarà assunto per lo sviluppo di questi strumenti. La supervisione sarà affidata a un esperto di database del gruppo di bioinformatica cui sarà anche affidato lo sviluppo dei database per i dati di genetica.

WP3 and WP4 : se gli utenti dovessero decidere di installare database per i dati di MicroArray e proteomica in un computer centrale gestito dal gruppo di Bioinformatica del CRS4, sarà necessario l'acquisto di un computer che funzioni da database server (8Gb or RAM, 6 TB of disk) e di una licenza per un database Oracle.

Il progetto si svolgerà entro la fine del dicembre 2007, e tutti gli sviluppi dovranno essere terminati entro tale data.

Durata e tempi di realizzazione del progetto

- Il progetto prenderà inizio il 15 settembre 2006 e finirà il 15 dicembre 2007 (15 mesi)
- Diagramma di GANTT del progetto

