

Titolo del progetto: Analisi di Big Data spaziali per la valutazione qualitativa di geohazards attraverso processi di *data mining*

Tutor: Dr Diego Di Martire

Co-tutor: Prof. Domenico Calcaterra

Programma di ricerca

La pluridecennale raccolta e archiviazione di dati geologici consentono oggi di accedere a nuove opportunità applicative nel campo delle Geoscienze, al punto che l'analisi dei grandi *datasets* disponibili può considerarsi il nuovo paradigma scientifico del XXI secolo. Secondo la stima del Comitato di Esperti delle Nazioni Unite per la gestione delle informazioni geospaziali globali (*United Nations Committee of Experts on Global Geospatial Information Management - UN-GGIM*), ogni giorno sono generati 2,5 quintilioni di byte di dati. Molte di queste informazioni sono oggi inserite e gestite in piattaforme GIS che ormai hanno raggiunto ottime prestazioni per le analisi matematiche e statistiche dei diversi livelli informativi (grazie allo sviluppo di specifici *tool*) e che sono in continua evoluzione. Inoltre, negli ultimi anni, i *big data* e l'intelligenza artificiale hanno fornito nuovi metodi e opportunità per molte applicazioni in Geologia (realizzazione di dettagliati modelli geologici del sottosuolo, geolocalizzazione di precisione con GPS, reti neurali artificiali associate al calcolo di suscettibilità da frana, tomografie sismiche 3D, individuazione e stima di georisorse, ecc.).

Tuttavia, i *big data* e le operazioni di *data mining* sono ancora agli inizi, i metodi e gli obiettivi non sono ancora del tutto consolidati, privi di un quadro teorico e applicativo unificato. Infatti, da un lato, le attuali tecniche di analisi dei *big data* possono essere trasversali, come la creazione di geodatabase relazionali e la gestione dei dati in tempo reale, d'altra parte, dovrebbero essere sviluppate tecniche specializzate per i *big data* spaziali basate sulla situazione reale, come l'indicizzazione spazio-temporale distribuita per i dati geologici, la valutazione dei tempi di ritorno di un fenomeno ai fini dell' pericolosità e l'uso di metodi di filtraggio che considerino la relazione spazio-temporale e la semantica geologica, peculiarità queste dello studioso in ambito geologico. Infatti, la maggior parte degli algoritmi ad oggi utilizzati per le valutazioni qualitative e quantitative nei vari settori di impiego dei *big data* si basano su tipologie di dati che non hanno proprietà spaziali e che quindi non interagiscono fisicamente tra loro e soprattutto con il contesto geologico, geomorfologico.

Le principali questioni scientifiche e tecnologiche riferite allo sviluppo delle Scienze Geologiche nell'era dei *big data* possono essere elencate come segue: (i) la gestione integrata dello *storage* e la gestione di dati strutturati, semi-strutturati e non strutturati, dati grandi e piccoli, dati ibridi e di precisione, (ii) modello e dati, modello di esplorazione statica e modello di monitoraggio dinamico, (iii) la combinazione di *data mining* e analisi dei dati, l'unificazione di correlazione e causalità, estrazione profonda e (iv) visualizzazione di *big data* delle Scienze Geologiche.

Il progetto di ricerca proposto si pone l'obiettivo di approfondire le tematiche sui processi di *data mining* a partire dall'analisi di *big data* fisici e spaziali al fine di caratterizzare, modellare e verificare l'occorrenza spazio-temporale di possibili tipologie di geohazards che interagiscono con il contesto geologico di taluni ambiti urbani ed extraurbani.

A tal fine, ci si avvarrà inizialmente dell'ingente mole di dati già disponibile per poi proseguire nella raccolta di dati specifici in base alle esigenze del progetto di ricerca; la conoscenza pregressa e le conoscenze acquisite durante il periodo di formazione consentiranno di progredire nello stato delle conoscenze, adottando metodologie aggiornate allo stato dell'arte. Nella prospettiva di un percorso verso le "città digitali smart", l'obiettivo finale del progetto di ricerca riguarderà lo sviluppo di piattaforme di supporto decisionale per la previsione e valutazione quantitativa nella gestione dei geohazard a partire dall'analisi di *datasets* geologici multi-sorgente, adeguatamente processati con algoritmi di analisi di *big data* opportunamente modellati.

Proposta per una posizione di dottorato

Sulla base della suddetta premessa scientifica, si propone una posizione di dottorato per il cui svolgimento saranno messe a frutto collaborazioni con centri di ricerca nazionali ed internazionali all'avanguardia sul tema dell'analisi di *big data* spazio-temporali (es.: University of Twente, British Geological Survey.....), ritenendo tali collaborazioni fondamentali per la crescita culturale del dottorando, oltre che per il raggiungimento di risultati scientifici condivisi e validati nell'ambito della comunità scientifica internazionale. Si ritiene, pertanto, che la ricerca possa essere articolata, nel corso dei tre anni di dottorato, così come di seguito descritto:

- primo anno: formazione istituzionalmente prevista per tutti i dottorandi del collegio e specifica per affrontare la tematica di ricerca; raccolta ed analisi ragionata della bibliografia specifica; acquisizione di competenze relative ai metodi di analisi e *data mining*; scelta delle aree di studio; raccolta di dati per la costruzione di un database specifico;
- secondo anno: integrazione delle conoscenze disponibili con acquisizione di dati *ad hoc* nelle aree di studio; messa a punto di una procedura di analisi multi-hazard inerente agli obiettivi del progetto ed alle aree studio individuate;
- terzo anno: verifica e validazione della procedura di analisi multi-hazard; redazione dell'elaborato di tesi.

Le potenziali aree di studio ricadranno in settori di territorio interessati da differenti geohazards noti in letteratura (frane, alluvioni...), dove è nota contemporaneamente la disponibilità di dati di monitoraggio.

Le spese per le attività di campo e per altre attività legate al progetto saranno coperte dai fondi dipartimentali del Gruppo di ricerca in Geologia Applicata e Geotecnica.

Il/la candidato/a che intende proporre la sua candidatura dovrà avere solide basi in analisi di rischi naturali, matematica, fisica, geo-statistica, conoscenze d'informatica, GIS. La conoscenza di tecniche e linguaggi di programmazione costituisce requisito preferenziale (Matlab, R, ...).