



(Geospatial)Data Management & Digital Innovation



Quest'opera è stata rilasciata con licenza Creative Commons
Attribuzione - Non commerciale - Non opere derivate 4.0 Internazionale.
Per leggere una copia della licenza visita il sito web
<http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Introduzione	3
Obiettivi specifici	4
Informazioni principali	5
Sede del Master	5
Target utenti	5
Sbocchi professionali	6
Iscrizioni	6
Percorso didattico	6
Skill	7
Moduli	8
Corsi	9
Docenti	11
Workshop/Eventi	11
Ottenere e visualizzare i dati. Geodata, Open Data e Big Data	11
Ninja Blender	11
PostgreSQL, funzioni avanzate ed esempi reali	11
La cartografia partecipata dati e strumenti a supporto della società civile e del mondo del lavoro	11
Copyright vs. Copyleft	12
Offensive Security	12
Wikipedia editathon	12
M'appare la Val di Cornia - OpenStreetmap party	12
Stage	12
Pubblicazioni	12
Risorse on-line	12
Syllabus	12

Introduzione

Il World Economic Forum prevede¹ che entro il 2020 quella del “data analyst” sarà tra le figure professionali più ricercate nelle aziende di tutto il mondo.

Anche in Italia l’interesse verso il mondo dei dati sta crescendo notevolmente tanto che, secondo una recente ricerca² realizzata da TAG Innovation School, in collaborazione con Cisco Italia e Intesa Sanpaolo, su un campione di 550 piccole e medie imprese italiane il 50% delle PMI assumerà un esperto di analisi dei dati entro i prossimi 3 anni.

Gli ultimi anni hanno visto un costante e diffuso sviluppo di nuove tecnologie, ampiamente utilizzate in molte discipline, scientifiche, mediche, umanistiche.

Le analisi urbane, territoriali, ambientali e del patrimonio e qualsiasi azione di studio e progettazione ad esse riferite richiedono una conoscenza dello stato di fatto di vari fattori fisici (localizzazione, dimensione, stato di conservazione, ecc.) e della loro variazione nel tempo.

L’analista di dati geospaziali richiesto oggi e nel prossimo futuro ha per obiettivo l’affrontare la complessità delle sfide del miglioramento della vivere delle comunità umane e nei relativi spazi urbano ambientale sociale che potrà essere monitorato e reso sostenibile attraverso un efficiente ed intelligente uso delle infrastrutture tecnologiche integrate.

Queste necessità si avvalgono anche di pratiche operative di rilievo metrico 3D e monitoraggio che possono usufruire: di dati già presenti e disponibili (cartografie, modelli 3D, banche dati georiferite); di dati appositamente acquisiti per colmare le lacune di informazione che derivano dall’analisi dei dati disponibili; di dati eventualmente reperibili con azioni di web retrieval e crowd sourcing.

L’integrazione di dati già disponibili richiede una fondamentale azione di generalizzazione e omogeneizzazione da un punto di vista geometrico e semantico che deve essere affrontata con rigore teorico e metodologico in quanto garantisce l’uso e la futura interoperabilità delle informazioni acquisite per le diverse analisi e azioni progettuali.

L’acquisizione di nuovi dati, a integrazione e/o completamento di quelli già esistenti, richiede l’utilizzo di sistemi e strumentazioni ai quali oggi si richiede basso impatto economico, sempre più orientati verso l’utilizzo di smart devices, come piattaforme UAS (Unmanned Aerial Systems) e sistemi robotici mobili terrestri. Tali sistemi devono essere adattati e utilizzati in modo tale da garantire il rapido inserimento delle informazioni raccolte all’interno delle strutture dati previste. Tutte le acquisizioni dati devono seguire protocolli operativi condivisi e definiti a priori in base alle caratteristiche di ogni sistema e di ogni tipologia di dato. A tal fine il complesso degli

¹ <https://www.weforum.org/agenda/2016/01/8-jobs-every-company-will-be-hiring-for-by-2020/>

² <https://www.wired.it/attualita/2017/06/23/business-data-analyst-infografica/>

standards legati all'informazione spaziale ed ai suoi processi di acquisizione manipolazione e comunicazione è sia altamente sviluppato, ma anche in ulteriore progressiva definizione e evoluzione.

In un'ottica di condivisione e integrazione dei dati di diversa natura e origine occorre sviluppare e definire le regole di formazione di sistemi GIS, WebGIS.

Oggi possiamo dire che questa rivoluzione tecnologica ha raggiunto una piena maturità, e il movimento Open Source ha sicuramente avuto un ruolo decisivo in questo sviluppo, dando la possibilità a chiunque di poter accedere a strumenti molto potenti, sia software che hardware.

Le nuove tecnologie ci offrono oggi gli strumenti più efficaci per veicolare informazioni, facilitandone la fruizione e costituiscono sempre di più il valore aggiunto di qualsiasi progetto, scientifico, culturale o di marketing.

Pensiamo alle possibilità oggi offerte dalla rete nella diffusione delle informazioni, o alla possibilità di raggiungere ovunque un pubblico sempre più vasto offrendo contenuti di qualità, sfruttando applicazioni e servizi esistenti.

Nell'ottica della condivisione e divulgazione della conoscenza e delle informazioni, le tecnologie Open Source rappresentano oggi gli strumenti migliori, in quanto nati e sviluppati proprio da una cultura di collaborazione e di scambio di conoscenze. Proprio questa è la strada tracciata, ad esempio, dalla recente modifica del Codice dell'Amministrazione Digitale, che incentiva l'utilizzo di soluzioni Open Source da parte delle Pubbliche Amministrazioni in alternativa a software di tipo proprietario.

Questo rapido sviluppo della tecnologia però, ha spostato l'attenzione degli utenti sugli "strumenti", trascurando l'elemento fondamentale, ovvero il "dato".

Conosciamo davvero i nostri dati? Sappiamo davvero come gestirli e come sfruttarli per i nostri scopi? Siamo capaci di manipolare dati grezzi, non strutturati, per ottenere nuove informazioni?

E' possibile rispondere a queste domande solo se possediamo una conoscenza approfondita della natura stessa dei dati che andremo a manipolare o che vogliamo creare.

Per questo la tecnologia ci viene sicuramente incontro, ma dobbiamo essere in grado di scegliere e, soprattutto, di utilizzare gli strumenti, tra i tanti a nostra disposizione, adatti ai nostri scopi.

Nell'era dei "BIG DATA" chi si appresta a lavorare con i dati, deve essere in grado di gestire l'insieme delle tecnologie e delle metodologie di analisi per estrapolare, analizzare e mettere in relazione un'enorme mole di dati eterogenei, strutturati e non, per scoprire i legami tra fenomeni diversi e prevedere quelli futuri.

Due aspetti fondamentali per chi lavora con i dati, e che verranno trattati in maniera approfondita, sono la loro conservazione nel tempo e la loro sicurezza.

Il master in "Data management & Digital Innovation" ha come obiettivo proprio quello di creare nuove figure professionali in grado di gestire il dato in ogni fase della sua vita, dalla

creazione, alla manipolazione fino alla fase finale, ovvero la condivisione, sfruttando al meglio la tecnologia a sua disposizione, senza tralasciare gli aspetti legislativi sulla libera circolazione dei dati e il fondamentale concetto di "sicurezza dei dati".

Obiettivi specifici

Alcune delle tecnologie che il professionista sarà in grado di usare sono:

acquisizioni dati tramite tecniche GPS/GNSS, topografia classica, LIDAR, fotogrammetria terrestre, anche da sensori low cost e sferici, fotogrammetria UAV, sistemi ibridi MMS (mobile mapping system) , modellazione 3D, elementi di cartografia e sistemi di riferimento cartografici, GIS e GIS 3D, Web GIS, database, analisi spaziali, analisi statistiche, web scraping, programmazione di webapp e realtà virtuale

Il Master si pone quattro obiettivi fondamentali:

1. formare professionisti nell'uso di tecnologie moderne, innovative e aperte, per gestire al meglio il percorso di recupero/creazione, analisi e pubblicazione di dati di natura diversa, dai set di dati limitati e strutturati, ai big data.
2. Alcune delle tecnologie che il professionista sarà in grado di usare sono:
 - 2.1 acquisizioni dati tramite tecniche GPS/GNSS
 - 2.2 topografia classica
 - 2.3 LIDAR
 - 2.4 fotogrammetria terrestre, anche da sensori low cost e sferici
 - 2.5 fotogrammetria UAV
 - 2.6 sistemi ibridi MMS (mobile mapping system)
 - 2.7 modellazione 3D
 - 2.8 elementi di cartografia e sistemi di riferimento cartografici
 - 2.9 GIS e GIS 3D
 - 2.10 Web GIS
 - 2.11 database
 - 2.12 analisi spaziali
 - 2.13 analisi statistiche
 - 2.14 web scraping
 - 2.15 programmazione di webapp e realtà virtuale
3. Creare competenze che permettano alle imprese di sfruttare i dati per creare nuovi modelli di business e generare vantaggio competitivo sul mercato
4. mostrare come non sia ormai più necessario ricorrere a tecnologie commerciali, che spesso hanno costi di licenza insostenibili per le aziende, e di come sia preferibile optare per soluzioni Open
5. rendere completamente autonomo il professionista nella pianificazione di progetti, anche complessi, che vanno dalla semplice creazione di un blog a lavori di business intelligence, fino alla gestione di un sistema informativo territoriale (SIT).

Al termine del Master ogni partecipante sarà in grado di:

- gestire un sistema linux
- comprende l'origine dei dati e le eventuali possibili anomalie
- analizzare il flusso informatico delle informazioni
- interpretare i dati con metodi statistici per identificare problemi di business
- sfruttare set di dati non strutturati già esistenti per creare nuove informazioni
- elaborare database per la gestione dei dati, compresi quelli spaziali
- creare e gestire Sistemi Informativi Geografici (GIS)
- elaborare cartografia tematica
- elaborare modelli 3D di oggetti e strutture utilizzando sia il rilievo strumentale che quello fotogrammetrico
- effettuare analisi statistiche su dati non strutturati
- elaborare sistemi web per la gestione e la pubblicazione online dei dati
- proteggere e rendere fruibili i propri dati sul web o su altri supporti

Informazioni principali

Il master universitario di I livello "Data management & Digital Innovation" è organizzato dalla società Parchi Val di Cornia in collaborazione con Arc-Team e con il Politecnico di Torino, con il patrocinio di

e con il sostegno di vari enti e associazioni quali i Comuni di Piombino, Campiglia Marittima, San Vincenzo, Suvereto e Sassetta.

Il master è nato grazie anche al sostegno di aziende e professionisti del settore, enti locali, centri di ricerca e università, con i quali forma un network di soggetti accomunati dalla consapevolezza che saper gestire i nostri dati è di fondamentale importanza nella nostra società.

Di seguito riportiamo l'elenco dei soggetti che hanno contribuito all'organizzazione del master:

....

Sede del Master

Le lezioni frontali e i workshop si svolgeranno nelle aule del Centro di Documentazione di Villa Lanzi, a Campiglia Marittima.

Sono previste esercitazioni in classe e fuori, nonché uscite in siti di importanza archeologica, storico-artistica o naturalistica, presenti nel territorio della Val di Cornia.

Target utenti

Fornendo una preparazione trasversale rispetto al proprio percorso di studi, il Master è rivolto a tutti quei laureati che vogliono approfondire la conoscenza di strumenti e tecniche di elaborazione dati sviluppati su piattaforme Open Source.

Il master permette anche, a chi lavora già con i dati, di aggiornare le proprie conoscenze su metodologie e tecniche di recente sviluppo, dando la possibilità di iscriversi ad un singolo modulo o ad un pacchetto di moduli (skill).

Non sono previsti particolari prerequisiti, eccetto le capacità di utilizzo base di un PC, che i masteristi sono tenuti ad avere e portare durante lo svolgimento delle lezioni, per l'installazione dei programmi utili ai fini del master e lo svolgimento delle esercitazioni pratiche.

Sbocchi professionali

Lo scopo del master è quello di formare figure professionali in grado di lavorare con qualunque tipo di dato, di conseguenza il ventaglio di possibilità lavorative è molto ampio e comprende una varietà di settori.

La figura del data analyst o data scientist è una figura chiave che può trovare un'occupazione presso enti pubblici, aziende private, istituti di ricerca e tutte quelle realtà lavorative che hanno bisogno al loro interno di professionalità dotate di una serie di competenze multidisciplinari, che siano in grado di gestire una grande mole di dati di natura eterogenea, organizzandoli e valorizzandoli per garantire efficienza e produrre nuova conoscenza.

Iscrizioni

Numero minimo iscritti: 20

Scadenza iscrizione: 20 Settembre 2019

Quota di partecipazione: € 3.500 + IVA 22% - Borse di studio disponibili

Requisiti di ammissione: per l'iscrizione al Master è necessaria una laurea di primo o secondo livello, o laurea vecchio ordinamento di qualunque facoltà.

L'utente che non vuole seguire l'intero master ma è interessato ad alcuni corsi, può scegliere se iscriversi al singolo modulo o al pacchetto di moduli (skills) che compongono una particolare area tematica (vd. capitolo "Percorso didattico")

Percorso didattico

Sono previste 620 ore di lezione e 350 ore di tirocinio formativo da svolgersi presso una società, un Ente pubblico o un centro di ricerca in cui poter sviluppare e applicare le tecnologie apprese.

Le lezioni si svolgeranno dal lunedì al venerdì.

Alla fine del periodo di stage verrà discusso il project work elaborato durante le ore di stage.

Lo schema seguente mostra il calendario provvisorio:

inizio corsi	?
fine corsi	?
inizio tirocinio	?
fine tirocinio	?
discussione tesi	?

La frequenza alle lezioni non è obbligatoria, ma al termine di ogni corso sono previste prove d'esame (orale, scritta o project work), il superamento delle quali è indispensabile per il conseguimento del titolo finale.

Skill

Il master è stato suddiviso in 5 aree tematiche (skill) ognuna delle quali si concentra su un aspetto fondamentale del percorso di apprendimento:

1. **Conoscenza base di linux**

Secondo il portale W3techs³, che si occupa di monitorare le principali tecnologie usate nel web, quasi il 70% dei server in rete gira su piattaforma linux. Oltre al web moltissimi enti di ricerca, banche, servizi di pubblica utilità hanno fatto una scelta di tipo "open". I motivi sono diversi ma possono essere riassunti in 2 aspetti fondamentali:

- dal punto di vista "tecnico" Linux garantisce una sicurezza maggiore, una diffusa e attiva comunità di sviluppatori e un controllo totale di ogni parte del sistema;
- da un punto di vista "economico" il beneficio per un'azienda o per un ente pubblico è totale, abbattendo completamente i costi di licenze, supporto e manutenzione.

³ https://w3techs.com/technologies/overview/operating_system/all

Questi sono alcuni dei motivi che ci hanno spinto ad utilizzare Linux come base di sviluppo per tutte le attività legate al master.

2. **Raccolta ed elaborazione dati**

I dati, per loro natura, sono eterogenei e chi lavora con essi deve essere in grado di gestire ogni tipologia di dato. Alla fine dei corsi che compongono questa skill, lo studente sarà in grado di:

- a. capire di che dati ha bisogno
- b. cercare dataset esistenti da inserire nel proprio modello statistico
- c. creare nuovi dataset, anche di tipo geografico o tridimensionale

I dati geografici e tridimensionali rappresentano il valore aggiunto del master poiché, a causa della loro natura complessa, sono di difficile gestione ma possono veicolare un altissimo livello di informazione. Essere in grado di gestire questi tipi di dati fornisce una competenza che può risultare di particolare interesse per le aziende.

3. **Organizzazione e gestione dei dati raccolti**

Quest'area tematica rappresenta il cuore di ogni progetto di gestione dati.

La creazione di una banca dati è un punto imprescindibile e di fondamentale importanza per diversi motivi, che vanno dalla sicurezza e la persistenza dei dati alla loro distribuzione, oltre ad avere una solida base per poter effettuare analisi statistiche e spaziali.

In quest'ottica il punto di riferimento assoluto è PostgreSQL e la sua estensione spaziale Postgis che, con altri software come R, GRASS e Qgis formano la più potente suite di programmi per fare analisi di qualsiasi tipo.

4. **Analisi dei dati**

I corsi di quest'area daranno allo studente le basi per effettuare diversi tipi di analisi sui dati raccolti. Lo studente imparerà a gestire vari formati di dati e apprenderà nozioni di statistica per effettuare analisi su dati, sia alfanumerici che spaziali.

5. **Pubblicazione e condivisione dei dati**

L'ultimo step della formazione tratterà un aspetto molto importante, ovvero la presentazione dei dati.

Se quando si parla di dati la "sostanza" è importante, altrettanto lo è la "forma": saper trasmettere le informazioni è fondamentale per una buona riuscita di qualsiasi progetto e, anche in questo caso, le moderne tecnologie ci aiutano.

Ogni analista di dati deve assolutamente conoscere anche un minimo di programmazione: durante i corsi che formano quest'area tematica, lo studente imparerà alcune tecniche di programmazione e le basi dei linguaggi più utilizzati (php, javascript, python), nonché come dare valore ai propri dati sui motori di

ricerca, entrando nel merito del funzionamento degli stessi, per fare ricerche di tipo statistico e valutazioni dei risultati.

Moduli

Le aree tematiche appena descritte sono suddivise in moduli, formati da uno o più corsi secondo lo schema seguente. Per il programma dettagliato di ogni singolo corso si rimanda al Syllabus presente alla fine del documento.

1. CONOSCENZA BASE DI LINUX
 - 1.1. Introduzione al mondo Open**
 - 1.1.1. open source e open data
 - 1.1.2. linux e i sistemi operativi aperti
 - 1.2. Lavorare con Linux**
 - 1.2.1. Installazione SO e gestione pacchetti
 - 1.2.2. La shell di linux
2. RACCOLTA ED ELABORAZIONE DATI
 - 2.1. Tecniche di rilievo strumentale bidimensionale**
 - 2.1.1. Rilievo fotogrammetrico
 - 2.2. Tecniche di rilievo strumentale tridimensionale**
 - 2.2.1. Structure From Motion, Image Based Modeling e altre tecniche di rilievo tridimensionale
 - 2.2.2. Modelli digitali 3D di oggetti reali: acquisizione, processing e utilizzo
 - 2.2.3. Modellazione 3D con Blender
 - 2.3. Ottenere dati dal web**
 - 2.3.1. Web scraping, data mining e altre tecniche per recuperare dati da database e servizi aperti
3. ORGANIZZAZIONE E GESTIONE DEI DATI RACCOLTI
 - 3.1. Database relazionali: progettazione, modellazione e manutenzione di una banca dati**
 - 3.1.1. Principi base di SQL per la progettazione e la modellazione di un database relazionale
 - 3.1.2. Come amministrare una database server
 - 3.2. Gestione dei dati spaziali**
 - 3.2.1. Nozioni fondamentali per lavorare con i dati spaziali: proiezioni e sistemi di riferimento
 - 3.2.2. Configurare un database server per la gestione del dato spaziale
 - 3.2.3. Geographic Information System (GIS)
4. ANALISI DEI DATI

4.1. Analisi statistiche dei dati raccolti

4.1.1. R, OpenRefine, Superset e altri strumenti per manipolare i nostri dati

4.2. Analisi spaziali

4.2.1. GRASS

5. PUBBLICAZIONE E CONDIVISIONE DEI DATI

5.1. Applicazioni web

5.1.1. I linguaggi di programmazione per il web

5.1.2. Creazione di una webapp per la gestione e la pubblicazione dei dati

5.2. Come proteggere i nostri dati

5.2.1. Tecniche per aumentare la sicurezza dei nostri dati

5.3. Come dare valore ai propri dati sul web

5.3.1. Funzionamento dei motori di ricerca e tecniche di analisi dei dati

Corsi

Di seguito una breve descrizione dei corsi riportati nel precedente paragrafo, per il programma dettagliato, gli argomenti d'esame e i docenti, si rimanda al Syllabus presente alla fine del documento.

1. open source e open data (8 ore)

Introduzione al mondo Open.

Verranno trattati tutti gli aspetti della filosofia "open" e le sue ripercussioni sul mondo del lavoro. Verranno sfatati alcuni miti legati all'Open Source e verrà mostrato come è possibile fare "business" anche con software aperti.

Verranno, inoltre, introdotte alcune nozioni tecniche le quali saranno trattate in maniera approfondita nel successivo corso

2. linux e i sistemi operativi aperti (8 ore)

Durante questa corso verranno spiegate le caratteristiche tecniche dei sistemi operativi aperti e le differenze tra essi, in modo da mettere in condizione lo studente di poter scegliere il sistema operativo più adatto ai propri scopi.

3. Installazione SO e gestione pacchetti (8 ore)

Lo studente verrà guidato in ogni fase dell'installazione di un sistema linux.

Lo studente in possesso di un proprio pc potrà scegliere se creare una partizione, in modo da non cancellare il sistema operativo già presente sul pc, oppure se dedicare l'intero hard-disk al nuovo sistema operativo.

Verranno mostrate, infine, le principali attività da svolgere per monitorare il proprio sistema, per installare, aggiornare o eliminare pacchetti di software

4. La shell di linux (8 ore)

La shell di linux rappresenta il cuore del sistema, lo strumento più potente

utilizzato da un utente linux, soprattutto nei casi in cui bisogna compiere operazioni su un sistema privo di interfaccia grafica come, ad esempio, un server.

5. Rilievo fotogrammetrico (32 ore)

Il rilievo fotogrammetrico costituisce una tecnica di documentazione che può raggiungere altissimi livelli di precisione rispetto al rilievo strumentale.

Dopo aver appreso le nozioni base di photomapping (fotogrammetria, tecniche di ripresa, sistemi di coordinate), verranno fatti una serie di esempi pratici

6. Structure From Motion, Image Based Modeling e altre tecniche di rilievo tridimensionale (40 ore)

Uno dei limiti principali ad un uso diffuso di strumenti di laser scanning nell'ambito dei Beni Culturali è dato dal costo elevato per l'acquisto delle strumentazioni necessarie e delle rispettive applicazioni, spesso nell'ordine delle decine di migliaia di euro. Tuttavia, oggi esistono tecniche e applicazioni Open Source derivate dalla Computer Vision (Structure from Motion e Image Based Modeling) che, partendo da semplici immagini digitali, consentono di realizzare rilievi tridimensionali paragonabili per qualità e accuratezza a quelli ottenuti con laser scanner

7. Modelli digitali 3D di oggetti reali: acquisizione, processing e utilizzo (40 ore)

8. Modellazione 3d con Blender (80 ore)

9. Web scraping, data mining e altre tecniche per recuperare dati da database e servizi aperti (16 ore)

10. Principi base di SQL per la progettazione e la modellazione di un database relazionale (80 ore)

11. Come amministrare una database server (16 ore)

12. Nozioni fondamentali per lavorare con i dati spaziali: proiezioni e sistemi di riferimento (8 ore)

13. Configurare un database server per la gestione del dato spaziale (24 ore)

14. Geographic Information System (GIS) (24 ore)

15. GRASS (40 ore)

16. R, OpenRefine, Superset e altri strumenti per manipolare i nostri dati (8 ore)

17. I linguaggi di programmazione per il web (60 ore)

18. Creazione di una webapp per la gestione e la pubblicazione dei dati (80 ore)

19. Tecniche per aumentare la sicurezza dei nostri dati (24 ore)

Docenti

Da definire

Workshop/Eventi

Ottenere e visualizzare i dati. Geodata, Open Data e Big Data

(Maurizio Napolitano)

Inserire descrizione e date

Ninja Blender

(Cicero Moraes)

Inserire descrizione e date

PostgreSQL, funzioni avanzate ed esempi reali

(2ndQuadrant)

Inserire descrizione e date

La cartografia partecipata dati e strumenti a supporto della società civile e del mondo del lavoro

(Andrea Borruso)

Inserire descrizione e date

Copyright vs. Copyleft

(Creative Commons Italia)

Inserire descrizione e date

Offensive Security

(E-Learn Security)

Inserire descrizione e date

Wikipedia editathon

(Wikimedia Italia)

Inserire descrizione e date

M'appare la Val di Cornia - OpenStreetmap party

(comunità italiana OSM)

Inserire descrizione e date

Stage

Pubblicazioni

Risorse on-line

Syllabus
