

## **Note relative alla costituzione di un CdS magistrale in Data Science**

*Carla Facchini*

*Presidente della Conferenza nazionale dei Direttori e responsabili didattici di area sociologica*

### **Mutamenti nella raccolta dei DATA e prospettive occupazionali nel settore**

Negli ultimi anni è cresciuta in modo esponenziale la disponibilità di dati raccolti sistematicamente e su larga scala, sia da parte di Enti pubblici, sia da parte di Enti privati ed aziende.

Per taluni di questi dati, si tratta di informazioni presenti anche precedentemente, ma il cui accesso è stato fortemente modificato dalle nuove tecnologie, per altri di materiale che solo le nuove strumentazioni/ tecnologie hanno permesso di raccogliere in modo così massiccio.

Tra i primi rientrano sia dati raccolti istituzionalmente –ad esempio i dati demografici (nati/matrimoni, morti, migrazioni), quelli sulle forze lavoro o quelle sul ricorso a strutture ospedaliere / prescrizioni farmaceutiche, sia materiali testuali di archivi pubblici e privati (si pensi alle sentenze dei Tribunali, ma anche ai giornali, molti dei quali hanno digitalizzato le annate precedenti facilitando in modo incredibile la ricerca accurata di documentazione pregressa).

Tra i dati di ‘nuova’ acquisizione vi sono invece quelli raccolti dalle diverse aziende a partire dai consumi dei propri clienti/utenti.

Comune a questi dati è la possibilità del tutto inedita di poter fruire e di poter elaborare in tempi brevi (potremmo ‘on demand’) e in modo quasi continuativo una massa enorme di informazioni, spesso colte nel loro divenire e quindi suscettibili di importanti letture longitudinali.

Tale nuova ricchezza di informazioni comporta la necessità di profili professionali in grado di affrontare e gestire una pluralità di questioni connesse alla raccolta, all’elaborazione, alla gestione e all’utilizzo di questi data base complessi e in continuo aggiornamento/ trasformazione.

Appare decisamente opportuna e interessante la proposta di costituzione di un corso di Laurea Magistrale in Data Science che abbia come obiettivo la formazione di nuove figure professionali in grado di trattare, analizzare ed interpretare la crescente disponibilità di dati digitali, ormai utilizzati in maniera consistente nella previsione e interpretazione di un vasto insieme di fenomeni sociali.

L’obiettivo formativo dovrà essere quello di formare competenze e capacità fortemente interdisciplinari nella Data Science per rispondere alle esigenze sia del mercato del lavoro, sia della ricerca più avanzata, attraverso la creazione di una figura professionale nuova, il Data Scientist, figura che sarà sempre più centrale nelle organizzazioni pubbliche e private che hanno l’esigenza di analizzare, per i propri obiettivi strategici e processi di sviluppo, dati e informazioni provenienti da un vasto insieme di fonti.

Al riguardo, basti citare che il Rapporto BigData@MIUR prevede almeno fino al 2020, una crescita dell’offerta di lavoro nei Big Data pari al 23% all’anno, contro il 19% di tutto il settore dell’Information Technology (IT) e il 6% globale.

D’altro canto, nei paesi sviluppati, il Data Scientist occupa posizioni professionali con remunerazioni che si collocano nella fascia alta delle categorie professionali IT e molte delle start up di successo nel mondo si collocano nell’area della Data Science.

I contesti occupazionali prevedibili possono quindi riguardare: Centri di ricerca e sviluppo, pubblici e privati (a partire dalla stessa Università;

- Centri ed aziende che operano nel settore del Data Providing and Data Brokerage,
- Pubblica amministrazione ed enti pubblici nazionali coinvolti nella gestione di grandi moli di dati (in primis INAIL, INPS, ISTAT)
- Enti locali (Regioni, grandi Comuni, loro Consorzi), ASL;

- Sedi formative, strutture socio-sanitarie;
- Aziende /Associazioni di aziende i cui prodotti siano rivolti ad un grande pubblico;
- Società di consulenze nei settori IT e Data processing.

## **Profilo professionale e competenze necessarie**

A fronte di tali prospettive occupazionali, queste nuove figure professionali dovranno possedere da un lato un'ottima capacità di predisporre la raccolta di tali dati, di elaborarli e di visualizzarli nei modi più opportuni, dall'altro le categorie concettuali e interpretative indispensabili per interpretare e dare un senso ai dati così raccolti, per evitare che, paradossalmente, proprio la loro sovrabbondanza rischi di comportarne una lettura banale e meramente descrittiva.

Da un lato occorrerà, quindi, che tali laureati abbiano una conoscenza approfondita dei principi teorici, metodologici e tecnici alla base della scienza dei dati; dall'altro competenze specifiche nelle scienze sociali - e specificamente nelle discipline sociologiche per la loro capacità di leggere fenomeni sociali e mutamenti in atto in base ad ipotesi teoriche complessive che suggeriscano non solo linee interpretative, ma la stessa scelta dei dati da raccogliere ed elaborare.

L'offerta formativa dovrà quindi essere fortemente interdisciplinare e ricomprendere, prioritariamente, le aree dell'informatica, della statistica e della sociologia.

Per quanto riguarda le competenze informatiche, il Data Scientist dovrà avere conoscenze approfondite oltre che dei fondamenti, degli aspetti applicativi dell'Informatica e del trattamento di grandi quantità di dati; dovrà inoltre saper analizzare e gestire i flussi di generazione, acquisizione, trasmissione e accesso ai dati; gestire e integrare archivi eterogenei di dati derivati da diversi processi/fonti; confrontare potenzialità e limiti dei diversi approcci computazionali.

Per quanto riguarda la formazione metodologica, occorrerà che i laureati abbiano:

- conoscenze approfondite del metodo scientifico di indagine e dell'utilizzo delle diverse metodologie applicate ai fenomeni sociali, economici, industriali;
- la capacità di individuare ed accedere alle fonti dei dati e di scegliere i metodi e i modelli più idonei ed efficaci, anche integrando tra loro le diverse metodologie di analisi dei dati, quantitativi e testuali; la capacità di utilizzare le tecniche più innovative utili per la modellizzazione delle rilevanze empiriche raccolte, compreso il campo delle regressioni speciali per lo studio di dati longitudinali e dello sviluppo di modelli predittivi.
- le competenze atte a sviluppare programmi di sviluppo mirati e flessibili, rapportandosi in modo propositivo ai processi decisionali e agli obiettivi dei decisori;

Nello stesso tempo, per quanto riguarda la capacità di contestualizzare e di interpretare i fenomeni di cui i dati raccolti tratteggiano gli andamenti, la cui rilevanza è rimarcata in esperienze internazionali di grande livello (si pensi, ad esempio, al *Master Degree in Social Data Science* dell'*Oxford Internet Institute*), è determinante non solo una formazione attenta agli aspetti epistemologici, ma anche all'acquisizione di quadri concettuali in grado di permettere la contestualizzazione dei dati raccolti nei diversi ambiti loro propri (occupazione, salute, utilizzo dei servizi educativi e socio/sanitari; comportamenti demografici, consumi e stili di vita, ecc. ). Così come è fondamentale sia una conoscenza approfondita dei processi (organizzativi, amministrativi e di mercato) che hanno prodotto i dati stessi, sia la capacità di cogliere le conseguenze (organizzative, metodologiche, sociali e politiche) che la raccolta di questi dati può comportare nei differenti contesti di ricerca, amministrativi e aziendali – in una logica di riflessività del proprio operare.

Oltre a queste competenze ‘specifiche’, occorrerà, infine, che i laureati acquisiscano capacità di tipo più trasversale, quali la capacità di presentare graficamente i risultati delle analisi/elaborazioni effettuate e di esporli in un linguaggio accessibile anche ad un vasto pubblico, e quella di imparare a lavorare in equipe, anche interdisciplinari.

A questo riguardo, può essere utile ricordare che un’indagine nazionale effettuata nel 2015 sui laureati di tutti i CdL in Sociologia ha evidenziato con forza da un lato come, per chi lavora nel campo della ricerca, vi sia un fortissimo apprezzamento sia delle competenze sociologiche specifiche (quali la capacità di interpretare i fenomeni sociali; di utilizzare – assieme- le diverse metodologie della ricerca, intrecciandole tra di loro, di predisporre e organizzare progetti di ricerca); sia di soft skills, quali, ad esempio, la capacità di presentare dati e relazioni in modo chiaro ed efficace e quella di lavorare in equipe in cui siano presenti professionalità diverse.

### **Destinatari del CdS magistrale in data Science**

Rispetto alle caratteristiche del profilo che si intende implementare, un Corso Magistrale in Data Science dovrebbe rivolgersi a tutti i laureati triennali interessati ad ampliare e sviluppare le proprie conoscenze verso le tecnologie informatiche, le tecniche statistiche e i quadri concettuali necessari per gestire, elaborare, analizzare e interpretare dati e informazioni provenienti da un vasto ed eterogeneo insieme di fonti. Laureati quindi provenienti sia da CdL triennali caratterizzati da competenze più tecniche (informatici, matematici o statistici), che da CdL connotati da competenze più volte alla contestualizzazione e alla interpretazione dei dati (economici o sociologici).

In un quadro di provenienze diversificate, sarebbe presumibilmente opportuno prevedere la possibilità di rafforzare assi formativi marcatamente carenti mediante o percorsi parzialmente diversificati, o la richiesta di assolvere debiti formativi.

Questo in modo da preparare sia esperti che operano nelle imprese che raccolgono i dati, sia ricercatori altamente qualificati in grado di padroneggiare al meglio questi dati in un’ottica più improntata alla ricerca scientifica.

### **Modalità didattiche**

Se è fondamentale una corretta calibratura dell’offerta formativa in termini di ambiti disciplinari, altrettanto rilevante dovrà essere l’attenzione alle modalità didattiche, dato che sono soprattutto tali modalità ad implementare le soft skills accennate precedentemente.

Credo sia abbastanza scontato che un corso di laurea di questo tipo debba prevedere una forte presenza di attività laboratoriali – in primis, ovviamente, quelli volti a padroneggiare i diversi software per la gestione, organizzazione e analisi dei big data, e alla sperimentazione e costruzione di tecniche e di modelli di analisi. Ma fondamentali sono anche attività laboratoriali mirate alla formulazione di disegni di ricerca e di innovazione digitale che tengano in considerazione i dati come costrutti socio-tecnici e in cui le dimensioni di piccolo gruppo consentano agli studenti di rapportarsi in modo interattivo non solo col docente, ma anche con i compagni di studio, in modo da imparare a lavorare con altri – e ancor più se con percorsi curriculari precedenti differenziati tra loro.

Ugualmente importanti sono tirocini e stage in imprese coinvolte/interessate a queste analisi – in modo che gli studenti possano mettersi alla prova da un punto di vista operativo e, nello stesso tempo, rapportarsi anche con i problemi quotidiani che la raccolta e la gestione dei dati comporta e con gli imprevisti che anche progetti ben organizzati e articolati difficilmente riescono ad evitare.