



**BIG**

**DATA**

**@MIUR**



**RAPPORTO**  
**DEL GRUPPO DI LAVORO**



---

## Executive summary

# 1

---

### Perché il MIUR si occupa di Big Data

- 1 – Cosa intendiamo per *Big Data*
- 2 – Cosa rende speciali i *Big Data*
- 3 – Una miniera di opportunità
- 4 – I *Big Data* nel mondo

# 2

---

### La formazione e i *Big Data*

- 1 – Cosa succede in Italia e negli altri Paesi
- 2 – Imparare i *Big Data* a scuola
- 3 – I *Big Data* nell'Alta Formazione:  
laurea, laurea magistrale, PhD, master
- 4 – I *Big Data* per le policy del MIUR:  
modifiche al sistema formativo,  
valutativo, di monitoraggio

# 3

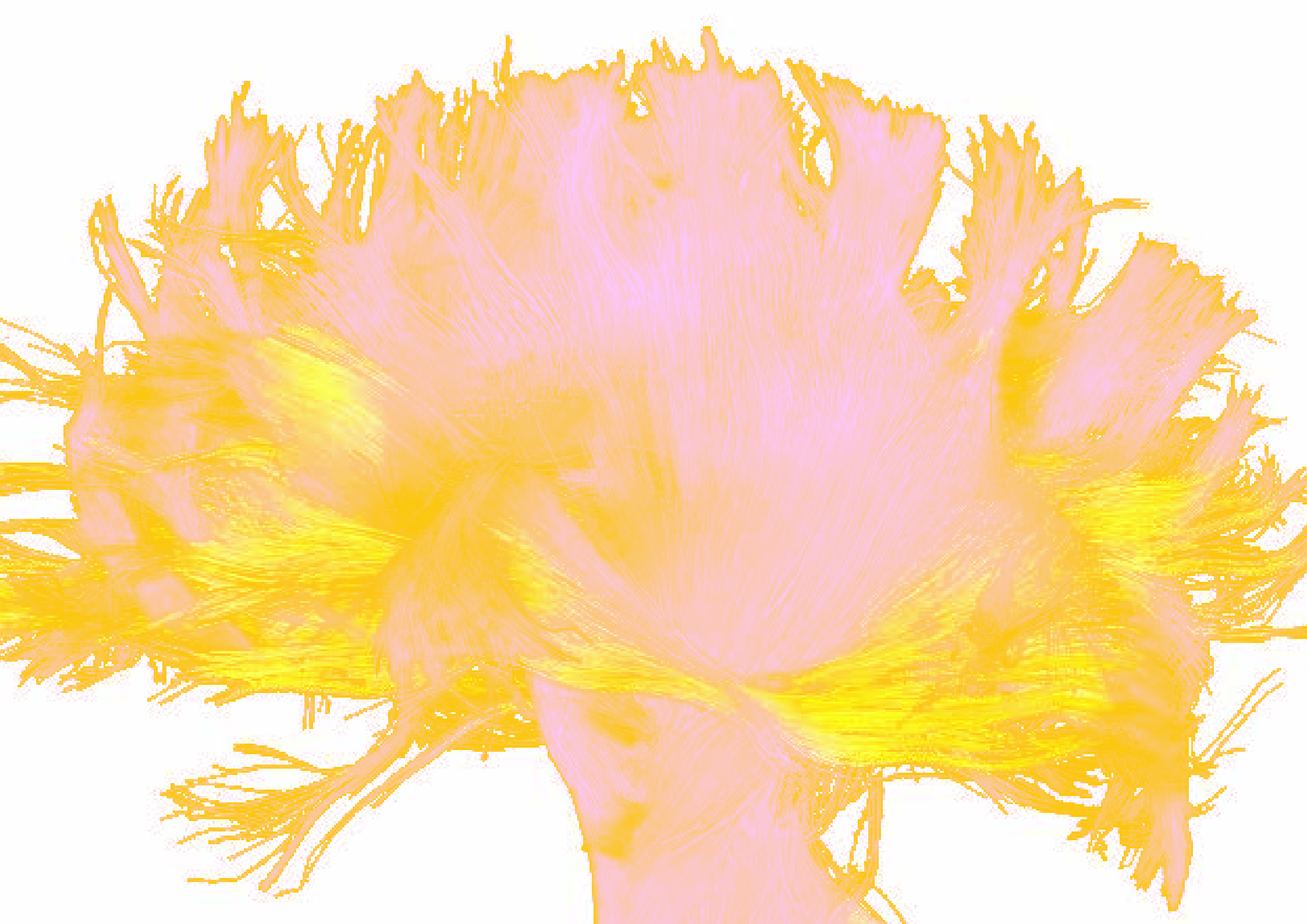
---

### Il MIUR e il patrimonio dei *Big Data*

- 1 – Le basi dei dati MIUR: il censimento
- 2 – Proposte
  - 2.1 – Policy sulla strutturazione e fruibilità dei dati MIUR
  - 2.2 – Policy per la valorizzazione dei dati pubblici e privati

---

### Riferimenti bibliografici



# Executive Summary

## il gruppo di lavoro sui *Big Data*

**Il Miur può contare su un vasto patrimonio di dati, da condividere in formato aperto per il riutilizzo: questo principio è già inserito nella legge 107/2015, La Buona Scuola. Tra le questioni più rilevanti per questo obiettivo:**

Questo documento fornisce i risultati del lavoro svolto dal gruppo istituito dal MIUR con il compito di analizzare policy e competenze a livello europeo e internazionale al fine di avanzare proposte per

- I. misure specifiche che il sistema formativo italiano dovrebbe adottare per essere linea con le migliori prassi internazionali,
- II. policy di sistema che il MIUR dovrebbe esprimere per valorizzare le potenzialità del fenomeno emergente dei *Big Data*.

## la formazione

L'analisi dell'offerta formativa nazionale e internazionale ha permesso di evidenziare alcune linee di azione per l'aggiornamento dei processi formativi, valutativi e di monitoraggio in un'ottica *Big Data*.

Si ritengono necessarie quindi azioni per:

- educare al "valore del dato" nello sviluppo delle competenze digitali fin dalla scuola primaria. Ciò avrà una necessaria ricaduta sull'adozione di metodologie didattiche innovative, con importanti opportunità sul monitoraggio e la valutazione degli stessi processi formativi;

- inserire corsi introduttivi alla *data science* in tutti i percorsi di laurea e post-laurea, a partire dalle triennali;
- incoraggiare percorsi specializzati sulla *data science* a livello di laurea magistrale e di dottorato di ricerca;
- sfruttare le potenzialità dei *Big Data* prodotti dagli ambienti di apprendimento digitali per il monitoraggio e la valutazione degli stessi processi formativi;
- incrementare le attività di formazione per il personale amministrativo ed economico-finanziario della scuola.

Il nostro sistema universitario presenta elementi di attenzione derivanti dalle rigidità disciplinari. L'urgenza di integrare le competenze delle diverse figure professionali oggi esistenti nell'attesa che i *digital native* si affaccino al mondo del lavoro, porta a suggerire la previsione di percorsi formativi mirati come passaggio obbligato: si sottolinea in questo contesto l'utilità di master di primo e secondo livello, preferibilmente soggetti ad accreditamento.

# accumulare e gestire i *Big Data*: policy e opportunità per la PA e per la ricerca

Possedere e avere accesso ai dati è un prerequisito per ogni azione in ambito *Big Data*, ma questa disponibilità non può essere data per scontata.

Per questo è necessario:

- stimolare, come azione di sistema, la creazione di basi di dati integrando informazioni provenienti dal pubblico e dal privato;
- definire standard (almeno) nazionali per il formato dei dati stessi;
- emanare bandi competitivi per il finanziamento di progetti di ricerca che abbiano i *Big Data* sia come metodo, sia come oggetto di studio;
- individuare risorse per il deposito e la conservazione di una mole considerevole di dati, nel quadro delle azioni del PNR sulle infrastrutture di ricerca elettroniche;
- premiare il riutilizzo e la citazione dei dati messi a disposizione dai ricercatori, tenendone conto nelle diverse fasi della valutazione dei ricercatori (ad es. VQR);
- incentivare forme di condivisione dei dati sia per i privati, sia per i soggetti operanti nel settore pubblico, anche prevedendo un piano di sensibilizzazione sul valore della condivisione con modalità che tutelino la proprietà intellettuale.

La condivisione dei dati, la loro integrazione e il loro riutilizzo producono valore, capacità di scelta e di innovazione che non saranno disponibili senza questa integrazione.

# i dati MIUR in una prospettiva Big Data

Il Miur può contare su un vasto patrimonio di dati che condivide in formato aperto per il riuso, principio inserito in legge 107/2015, La Buona Scuola. Per valorizzarlo occorre adoperarsi perchè siano realizzati i punti di seguito:

- un'analisi dettagliata del vasto patrimonio di dati gestito direttamente dal MIUR rivela che un grande potenziale di sviluppo potrà derivare da una più stretta integrazione di questi dati che sono nati disaggregati per motivi storici e istituzionali;
- un vero salto qualitativo sarà reso possibile solo tramite l'integrazione di questi dati con quelli disponibili presso altri soggetti pubblici e privati;
- nel rapporto sono delineati alcuni casi e sono proposte delle linee di azione perché il MIUR si doti di nuovi strumenti per l'elaborazione di policy basate sull'analisi di queste informazioni integrate, ma non solo: questo patrimonio di dati integrati, attraverso il loro riuso in contesti anche lontani da quelli che ne hanno motivato la registrazione, rappresenta uno strumento determinante per l'innovazione in diversi settori della ricerca con importanti ricadute anche sul tessuto produttivo.

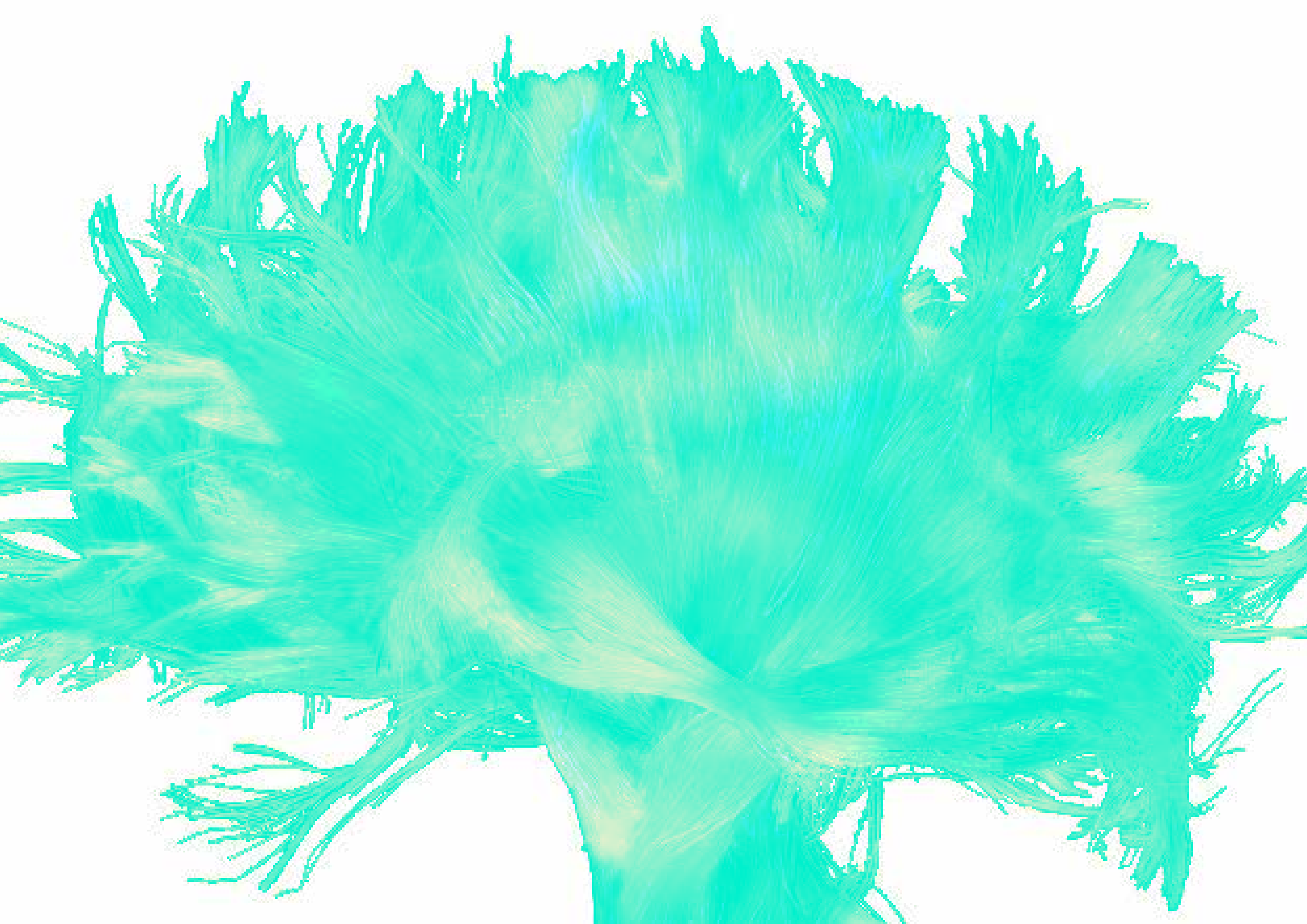




# conclusioni

Questo rapporto delinea un campo emergente di cui si vedono già le prime ricadute a livello globale, ma soprattutto appare il chiaro potenziale per un impatto di straordinario rilievo e grandissima pervasività. L'Italia è nelle condizioni di assumere un ruolo competitivo nello scenario internazionale, ma i diversi attori pubblici e privati devono agire con tempestività. È convinzione del gruppo che una rapida attuazione delle proposte qui delineate consentirà al MIUR non solo di beneficiare degli sviluppi offerti dalla *Data Science* nello svolgimento della propria funzione, ma anche di proporre un modello di *best practice* che potrà avere un prezioso effetto traino su altri settori della pubblica amministrazione, sul mondo della ricerca e dell'innovazione con importanti ricadute sul sistema produttivo.





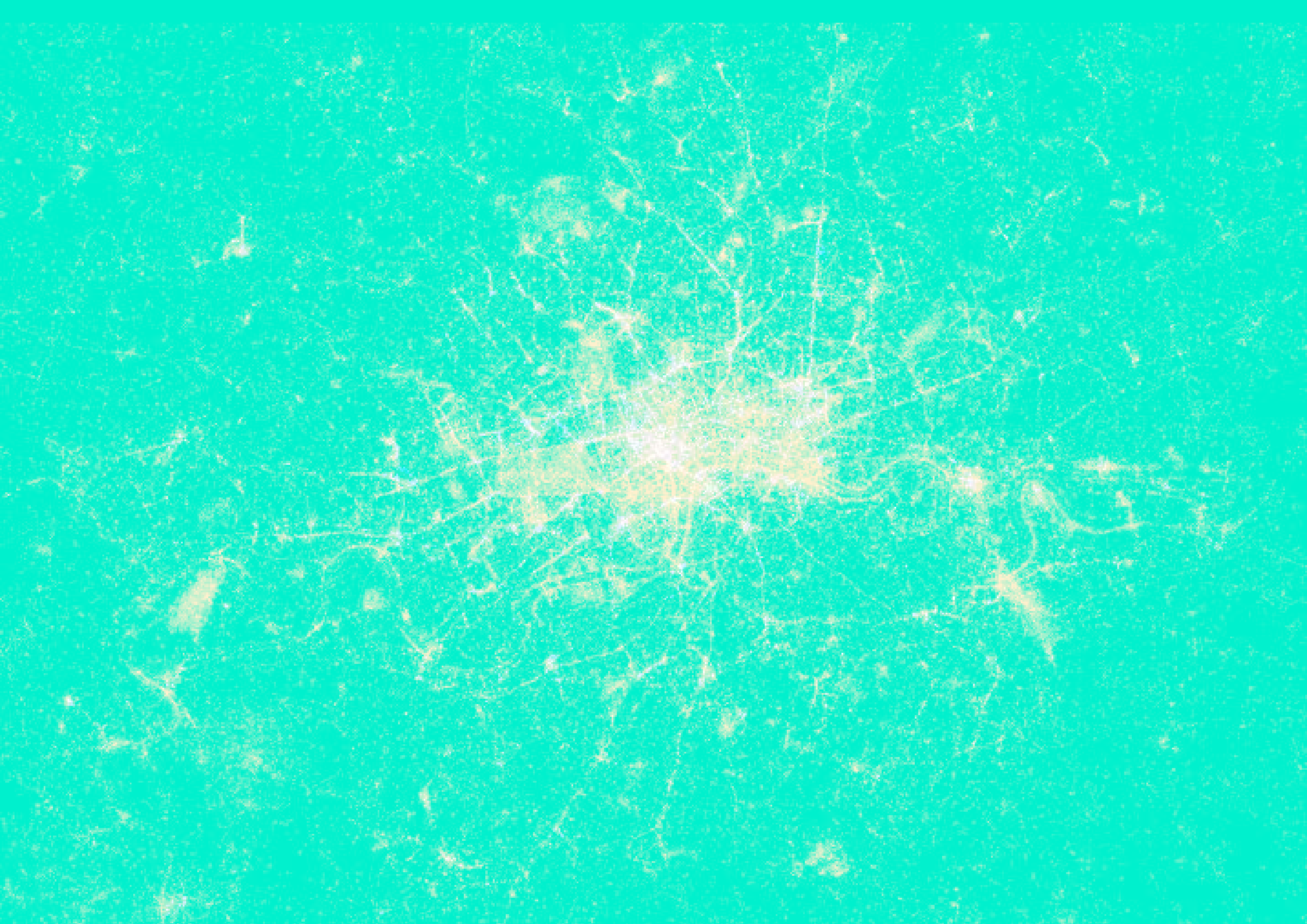
**1**

**—**

**I N T R O —**

**D U —**

**Z I O N E**



# Perché il MIUR si occupa di *Big Data*

Il gruppo di lavoro sui *Big Data* è stato insediato al MIUR con il compito di:

- mappare, in chiave anche comparata a livello europeo e internazionale, i principali centri, universitari e non, che operano nel settore dei *Big Data* con riferimento alla formazione e costruzione di competenze;
- identificare le possibili misure che il sistema formativo italiano dovrebbe adottare per essere in linea con le migliori prassi internazionali valorizzando le proprie specificità e, ove possibile, assumere un ruolo guida in questo settore in rapida evoluzione;
- effettuare la ricognizione delle banche dati del Ministero, relative al settore dell'istruzione scolastica, dell'università e dell'alta formazione artistica e musicale, nonché della ricerca scientifica, e proporre azioni che il Ministero potrebbe intraprendere per valorizzare le stesse e renderle strumento utile all'elaborazione di scelte di policy strategiche.

Questo rapporto rappresenta la sintesi dei risultati raccolti dal gruppo di lavoro, è stato condiviso dall'insieme dei suoi membri ed è stato redatto dopo una serie di consultazioni e richieste di documenti a diversi esperti del mondo accademico, della ricerca, dell'amministrazione pubblica e dell'impresa.

I componenti del gruppo ringraziano sentitamente tutti coloro che hanno contribuito con informazioni, idee e suggerimenti e, in particolare, Maria Teresa Sagri, Elisa Letizia, Giordano Mancini e Daniele Regoli che hanno affiancato il gruppo in tutti i suoi lavori.

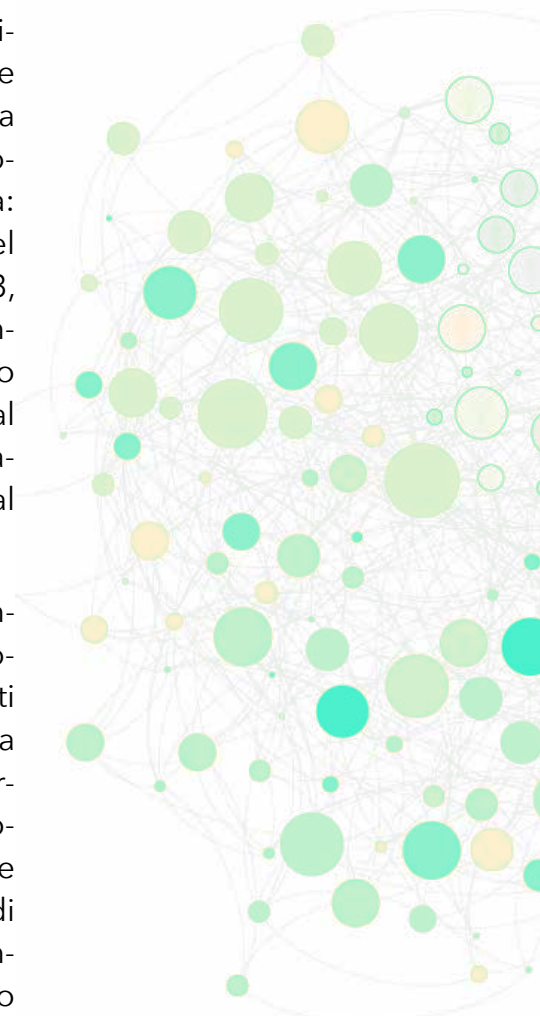
# 1. Cosa intendiamo per Big Data

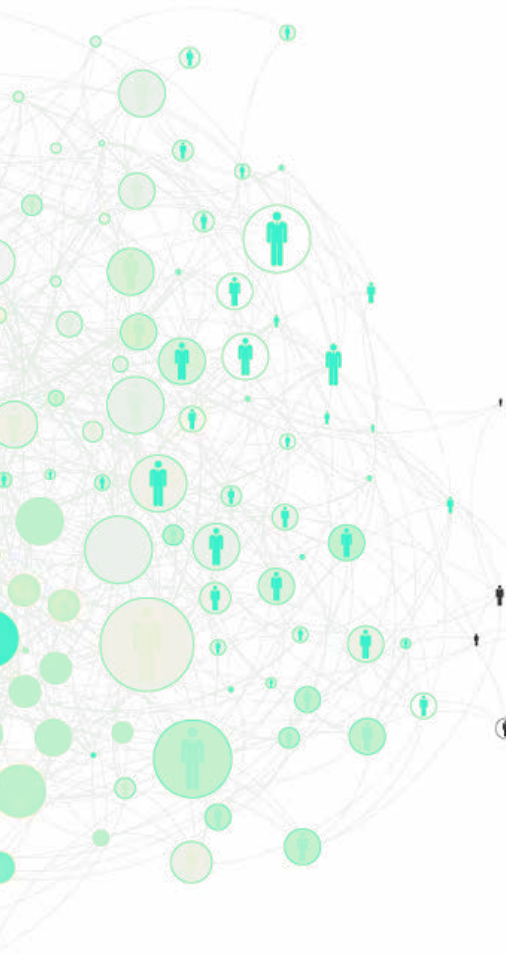
Ogni telefono cellulare durante il suo funzionamento registra continuamente una grande quantità di informazioni sul luogo in cui si trova, sulla velocità dei suoi spostamenti, sulla cella più vicina alla quale si può collegare, sugli altri telefoni con i quali comunica, sui servizi applicativi che gli utenti adottano per svolgere attività bancarie, leggere le informazioni, cercare biglietti per i loro viaggi o per andare al cinema, e così via. Milioni di automobili raccolgono dati sui percorsi, le soste, gli stili di guida, lo stato del motore e dei freni con i sensori, i computer di bordo, le scatole nere che registrano informazioni in ogni momento. Ogni contatore intelligente dell'energia elettrica, ogni orologio che raccoglie dati sull'attività sportiva svolta dal suo proprietario, ogni macchina agricola, ogni termostato moderno, raccoglie e comunica dati. Nelle fabbriche ogni macchina è dotata di sensori che registrano e comunicano informazioni, per migliorare i prodotti, per la manutenzione, per l'efficientamento della produzione. Nei campi i sensori avvertono i contadini dei cambiamenti climatici o della necessità di annaffiatura. Dall'alto, migliaia di satelliti registrano ogni cosa con

precisione crescente, mentre servono miliardi di persone con informazioni sulla loro localizzazione, che a loro volta vengono registrate. Inesorabilmente, dall'inizio del nuovo millennio, un cambiamento radicale è avvenuto nel modo con il quale l'umanità registra ciò che sa. Ne offre un'idea Martin Hilbert, ricercatore alla scuola Annenberg dell'università di California: nel 2000, il 25% di tutta l'informazione prodotta nel mondo era registrato su supporto digitale; nel 2013, il 98% era digitale. Si tratta di un fenomeno emergente: secondo IBM, il 90% dei dati digitali globali è stato generato negli ultimi due anni [1] e il McKinsey Global Institute ha stimato che il volume di dati immagazzinati su scala mondiale crescerà del 40% all'anno fino al 2020 [2].

Il numero di oggetti digitali connessi cresce esponenzialmente, con il crollo del costo e l'aumento del valore di connettere, memorizzare ed elaborare dati. Tutti questi dati nascono funzionali a uno scopo, ma una volta raccolti possono essere usati per mille altre forme di analisi, correlazione, modellizzazione. Le conoscenze che ne derivano non richiedono più modelli e teorie di riferimento, ma nascono dall'osservazione di fatti registrati dalla popolazione analizzata nel momento stesso in cui questa svolge le attività che si vogliono comprendere.

Questi sono i *Big Data*: non sono solo grandi quantità di dati. Ma dati funzionali alle necessità operative di miliardi di oggetti che registrano i comportamenti di persone e macchine in modo capillare che diventano





una piattaforma di informazioni capace di alimentare diversissime forme di generazione di conoscenza. Possono servire al miglioramento dei servizi cittadini o alla sicurezza delle strade, ma anche alla sorveglianza della popolazione o al miglioramento dell'efficacia della pubblicità. Sicché, tutte queste conoscenze informano sulla vita privata delle persone con un'intrusività mai conosciuta prima. Dal punto di vista dei diritti umani, questa è una sfida da vincere. Le opportunità vanno colte e i rischi conosciuti. Anche perché questi dati sono immagazzinati in *cloud*, cioè su computer remoti connessi a internet raccolti e gestiti in grandi *datacenter* spesso localizzati fuori dall'Europa e quindi governati secondo le normative di altri paesi, Stati Uniti in testa.

Tecnicamente, questi grandi insiemi di dati sono raccolti in *dataset* di grandi dimensioni che possono essere gestiti, conservati e analizzati solo facendo uso di tecnologie avanzate. Un'indicazione quantitativa in materia può essere ricavata dalla proiezione sulla domanda di spazio di memoria per l'immagazzinamento di questi dati che, per il 2017, viene dimensionata in 16 Zettabyte (ovvero 16 mila miliardi di Gigabyte). Una taglia senza precedenti che pone seri problemi di sostenibilità e affidabilità delle soluzioni oggi disponibili per la conservazione dei dati nel lungo periodo (cioè alcune decine di anni) e rappresenta una delle principali sfide tecnologiche del futuro. Le frontiere sono nell'evoluzione della connessione mobile e nella diffusione di oggetti dotati di sensori e connettività (*Internet of Things*). Ci si aspetta che nei prossimi anni crescerà

ancora esponenzialmente il numero e la diversità delle possibili sorgenti di dati, in una complessità alimentata da un processo di coalescenza per cui insiemi di dati un tempo fisicamente o logicamente isolati **sono ora accessibili insieme**.

Per l'Italia, secondo paese manifatturiero europeo, si tratta di una sfida da non perdere: *l'Industria 4.0*, la nuova generazione di imprese che sfruttano le tecnologie dell'*Internet of Things*, del *cloud computing* e dei *Big Data*, con sistemi e prodotti sempre più interconnessi e in comunicazione continua, dimostra come la rete e i dati stiano cambiando il modo stesso di fare impresa.

Epistemologicamente, i *Big Data* cambiano la prospettiva: da *top down* a *bottom up*. La *data science* è orientata a estrarre informazione e valore dai dati analizzandoli secondo schemi di indagine diversi da quelli per i quali sono stati raccolti, rispondendo alle domande attraverso il riconoscimento di modelli emergenti dal comportamento dei dati, usando strumenti e algoritmi e forme di visualizzazione in pieno sviluppo. Generando forme di innovazione guidata dai dati.

## 2. Cosa rende speciali i *Big Data*

I *Big Data* sono intrinsecamente diversi dai normali *dataset* raccolti nei *database* tradizionali. Vi è un sostanziale accordo nell'evidenziare quattro caratteristiche alla base di questa diversità:

i. **volume.** La dimensione è una caratteristica particolarmente importante per i dati generati da sistemi automatizzati; ad esempio i sensori installati in un singolo motore a jet generano circa 20 TB di dati ogni ora, che moltiplicato per il numero di aerei di linea in volo ogni giorno (circa 25000) implica la generazione di diversi Petabyte, PB, di dati ogni giorno.

ii. **velocità.** Sebbene il volume di dati provenienti da *social network* sia più ridotto rispetto ai *machine-generated data*, la grande velocità con cui possono essere prodotti genera comunque grandi volumi di dati (ad esempio si stima che nel 2011 il solo Twitter generasse circa 8 TB di dati ogni giorno; questo valore nel 2015 si aggira sui 12 TB).

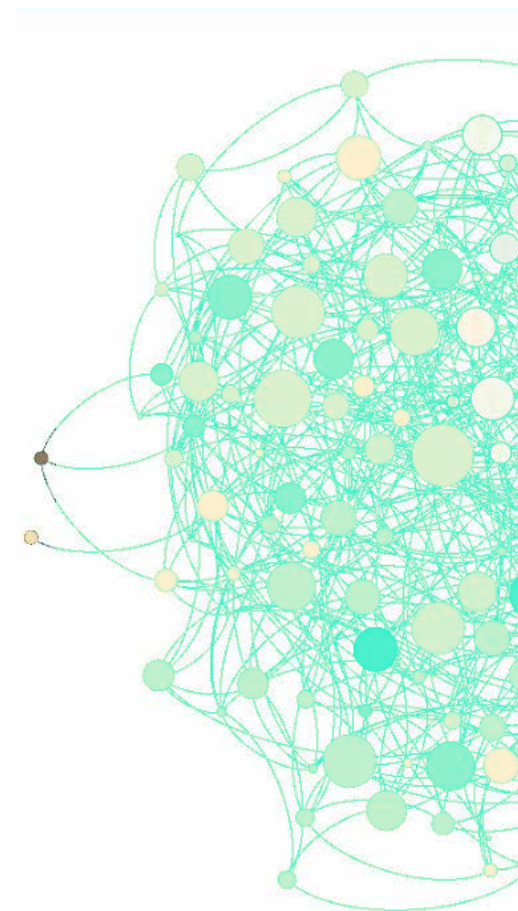
iii. **varietà.** Nei *dataset* tradizionali l'informazione è solitamente generata in modo omogeneo da un numero ristretto di sorgenti e secondo uno schema stabilito. Nei *Big Data*, invece la varietà delle sorgenti è eterogenea e in continua evoluzione.

iv. **valore.** Rispetto ai *dataset* tradizionali, l'informazione utile è qui nascosta dalla quantità di dati. La sfida dunque è individuare ed estrarre quello che può creare valore.

## 3. Una miniera di opportunità

"*Data is the new oil*": alla base di questa affermazione c'è l'idea secondo la quale la quantità di informazioni racchiuse nei *Big Data* è tale da permettere di fare previsioni (e quindi elaborare strategie) basate sull'osservazione di correlazioni tra e nei **dati reali** invece che su un'estrapolazione basata su modelli a priori (si parla di *data-driven decision*). Poiché una delle caratteristiche dei *Big Data* è la velocità con cui vengono generati, ci si aspetta che - avendo le necessarie capacità - sia possibile effettuare previsioni in tempi molto rapidi, anche senza attendere che i dati siano consolidati (*now-casting*) in modo da anticipare o intercettare immediatamente cambiamenti economici o sociali in atto.

La vastità delle opportunità offerte dai *Big Data* è tale che anche le più ambiziose pianificazioni non riescono ad anticipare tutte le innovazioni che ne potranno derivare. Per le imprese private questa è una sfida. Per il settore pubblico apre a considerazioni strategiche: una grande disponibilità di dati pubblici (aperti), uti-





lizzabili dalle imprese consolidate e dalle *startup*, oltre che dalla ricerca universitaria, diventa essa stessa un'**infrastruttura capace di generare innovazione** potenzialmente importante. E tutto questo può a sua volta produrre una condizione di maggiore efficienza per tutto il sistema dei servizi pubblici.

Oltre che un fenomeno complesso i *Big Data* sono quindi una **risorsa strategica** fondamentale per lo sviluppo, ma l'accesso a tale risorsa richiede la corretta gestione di:

- 1. raccolta e conservazione dei dati:** come far sì che i dati esistenti in vari settori (oggi non raccolti oppure frammentati) diventino *Big Data*
- 2. analisi ed estrazione dell'informazione:** come formare figure e porre le condizioni per lo sviluppo di metodi in grado di analizzare i dati
- 3. data driven decision:** come ottenere una ricaduta positiva sulla collettività (o sull'azienda) partendo dalle informazioni estratte.

Da questo nasce la necessità di politiche, di indirizzi che in questo momento vengono elaborati a livello globale.

## 4. I Big Data nel mondo

Negli ultimi anni, vista la diffusione e le opportunità legate ai *Big Data*, i governi delle economie più avanzate hanno avviato una discussione sulle politiche e gli impegni da adottare.

### **UNIONE EUROPEA**

Documento di riferimento:

*Towards a thriving data-driven economy, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, 2014 [5]*

"Nel Luglio 2014, La Commissione ha delineato una nuova strategia sui *Big Data*, per supportare e incentivare la transizione verso una *data-driven economy* in Europa. L'economia data-driven stimolerà ricerca e innovazione sui dati e allo stesso tempo porterà a nuove opportunità di business e a nuove disponibilità di competenze e capitali in tutta Europa, in modo particolare per le PMI"

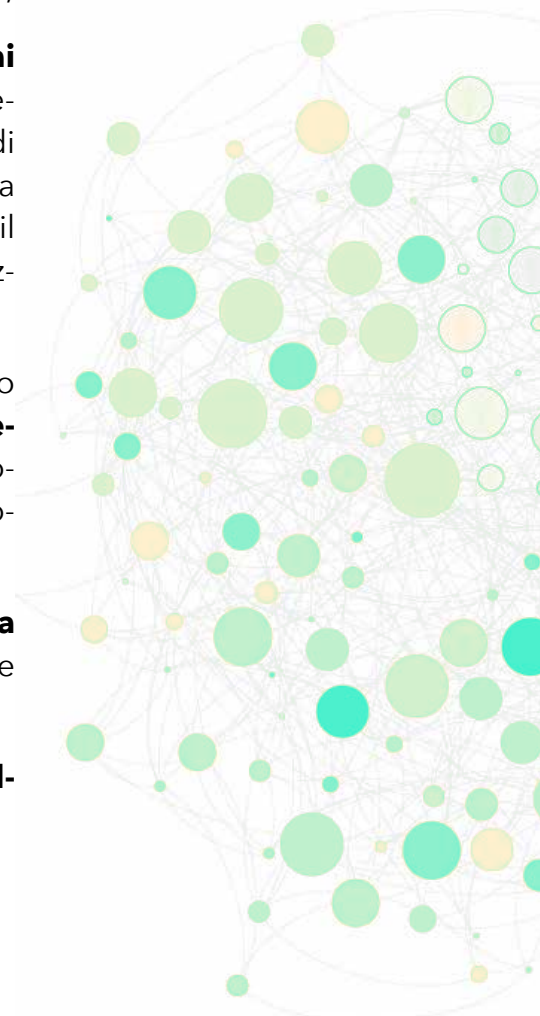
La Commissione sottolinea come l'Europa sia in forte ritardo rispetto agli Stati Uniti, sia dal punto di vista

delle politiche governative sia, in particolar modo, per quel che riguarda le risorse e le tecnologie industriali. La ricerca pubblica e privata su metodi e tecnologie per *Big Data* nell'Unione Europea è giudicata insufficiente e le attività corrispondenti in larga misura prive di coordinamento. C'è inoltre una mancanza di figure competenti che traducano gli sviluppi tecnologici in concrete opportunità di business [4]. Infine, la complessità dell'aspetto legale e le difficoltà ad accedere ai dati creano forti barriere di ingresso per le imprese.

Nel documento sono evidenziati alcuni punti chiave che la Commissione ritiene necessario adottare per colmare questo ritardo:

- **sostenere iniziative "faro"** (*lighthouse initiatives*) sui dati, capaci di accrescere la competitività e la qualità dei servizi pubblici e della vita dei cittadini. Iniziative "faro" massimizzano l'impatto dei finanziamenti dell'Unione all'interno di settori strategici per l'economia europea. Alcuni esempi sono: sanità (medicina personalizzata), gestione integrata di trasporti e logistica per intere regioni, gestione della filiera alimentare tracciando i prodotti dal produttore al consumatore, ecc...;
- **sviluppare le tecnologie**, le infrastrutture sottostanti e le competenze necessarie, in particolare a beneficio delle PMI;
- condividere, utilizzare e sviluppare in larga scala le proprie **risorse di dati pubblici** e infrastrutture di dati per la ricerca;

- **focalizzare la ricerca pubblica** sulla tecnologia, sull'aspetto legale e su altri possibili punti deboli;
- assicurarsi che il **quadro giuridico relativo ai dati**, come ad esempio questioni di interoperabilità, protezione dei dati, sicurezza, diritti di proprietà intellettuale, sia *data-friendly*, porti a maggiori certezze in fatto di regolamenti per il business e crei fiducia nei consumatori nell'utilizzo dei dati e delle tecnologie associate;
- portare a compimento in tempi brevi il processo legislativo dell'Unione sulla riforma della **protezione dei dati** e sulla sicurezza dell'informazione, e sostenere a questo scopo il dialogo e la cooperazione tra le autorità di controllo;
- accelerare la **digitalizzazione della pubblica amministrazione** e dei servizi aumentandone l'efficienza;
- incentivare appalti pubblici per **trasferire i risultati della tecnologia dei dati sul mercato**.



## REGNO UNITO

Nel documento *Big Data Policies, 2014* [7] redatto all'interno del progetto europeo BYTE (*Big Data Roadmap and cross-disciplinary community for addressing societal externalities*) si legge:

“Le politiche sui *Big Data* variano molto tra gli Stati Membri. Il Regno Unito è un aprifila in tutti gli aspetti riguardanti i *Big Data*. Anche se non si è dotato di un preciso documento di politiche, tutti gli elementi di discussione all'interno dell'Unione Europea sono stati in qualche modo anticipati o influenzati da discussioni nel Regno Unito.”

Il 9 Ottobre 2013, con un finanziamento di 600 milioni di Sterline, il governo del Regno Unito ha pubblicato le *Eight Great Technologies* [8], dove *Big Data* figura come prima voce. In quella stessa occasione, 42 milioni di Sterline sono state stanziare per il finanziamento dei primi 5 anni di un nuovo istituto di ricerca sui *Big Data* e Data Science, l'Alan Turing Institute, con ulteriori finanziamenti da 5 università britanniche (Cambridge, Edinburgh, Oxford, UCL, Warwick) e alcuni partner industriali. L'Alan Turing Institute ha cominciato le attività nel 2015, è un istituto di ricerca di diritto privato e si propone come “l'istituto nazionale britannico per la Data Science”.

Nel documento governativo *Seizing the Data Opportunities, 2013* [9] si legge:

“Questa strategia si concentra su tre aspetti globali. La prima è il capitale umano, una forza lavoro preparata e dei cittadini fiduciosi sulle opportunità offerte dai dati. La seconda riguarda gli strumenti e le infrastrutture disponibili per raccogliere ed analizzare i dati. La terza è la stessa disponibilità dei dati, le opportunità nell'uso dei dati sono subordinate alla possibilità da parte di consumatori, imprese e mondo accademico di accedere e condividere opportunamente i dati.”

Ulteriore aspetto di rilievo è il ruolo importante del Regno Unito come *repository* di basi di dati di rilievo per la *Big Science* in diversi ambiti, ma con un fuoco per i dati di interesse biomedico (SKA, 100000 Genomes, Elixir).

## FRANCIA

Nel 2013 il governo francese ha lanciato il piano *La Nouvelle France Industrielle*, un insieme di 34 piani industriali, uno dei quali specifico sui *Big Data*. Nel 2015 il governo ha poi lanciato la Fase 2 del piano, comprendente questa volta 9 punti chiave [10], di cui uno è *l'Économie des données* e che fa riferimento esplicito al valore insito nell'utilizzo opportuno dei *Big Data*.

In particolare il documento *Big Data: Feuille De Route, 2014* [11] è piuttosto articolato, con proposte di azioni concrete a breve e lungo termine.

Prendendo in esame il caso della formazione, guidata dal Ministero dell'istruzione francese (MENESR), il documento punta su azioni a breve termine, come la creazione di corsi MOOC dedicati alla *Data Science* e l'incoraggiamento dell'insegnamento della statistica nelle scuole, e a lungo termine, come la valutazione dei bisogni del mercato del lavoro e la creazione di corsi universitari dedicati alla figura del *Data Scientist*.

Concludendo, con le parole del documento *Big Data Policies* del progetto BYTE [8]:

"In generale, questa rassegna mostra che le politiche di *Big Data* si trovano decisamente in una fase di sviluppo, sia in Europa che nel resto del mondo. Se alcuni paesi e alcuni settori sono più avanti di altri, nessun contesto specifico è giunto ad un'esauriva politica sui *Big Data* che prenda in considerazione tutti gli aspetti rilevanti qui identificati".

## STATI UNITI D'AMERICA

Documenti di riferimento:

*Office of Science and Technology Policy (OSTP) -- Executive Office of the President: Big Data Initiative Release, 29.03.2012 [7] e Seizing Opportunities, Preserving Values (Maggio 2014), redatto da un gruppo di lavoro guidato dal Counselor of the President John Podesta.*

È imminente la pubblicazione di un documento di policy aggiornato.

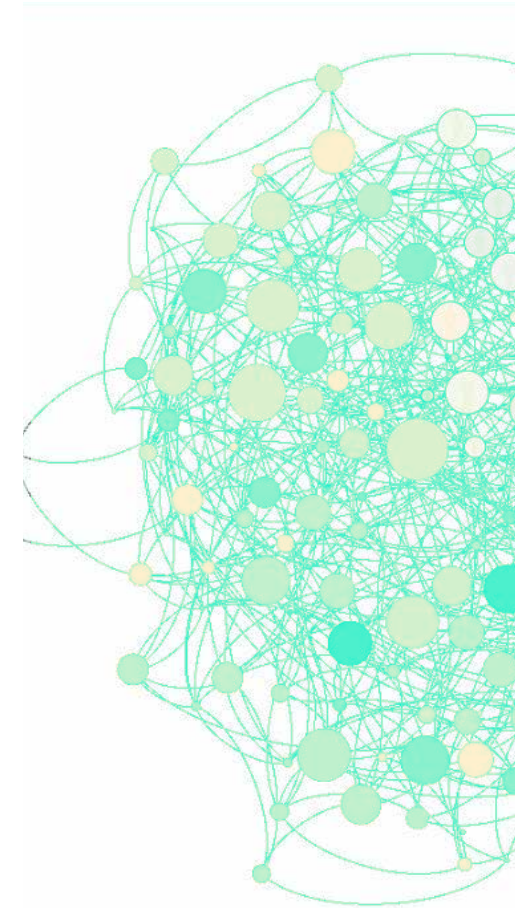
Nel marzo 2012, l'Amministrazione Obama ha annunciato la *Big Data Research and Development Initiative*:

"Migliorando la nostra capacità di estrarre conoscenza e comprensione da grandi e complessi insiemi di dati digitali, l'iniziativa si propone di aiutare a risolvere alcune delle sfide più incalzanti per la Nazione".

Lanciando l'iniziativa, sei tra Dipartimenti e Agenzie federali hanno annunciato un finanziamento di più di 200 milioni di dollari in nuovi progetti che mirano a migliorare in modo sostanziale le tecnologie e i metodi necessari per accedere ed organizzare grandi volumi di dati, ed estrarne informazioni.

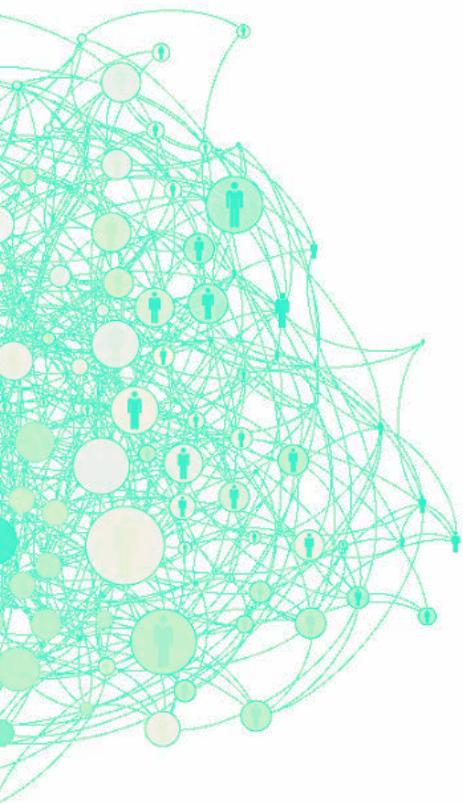
"L'OSTP di concerto con diversi Dipartimenti e Agenzie federali, ha creato la *Big Data Research and Development Initiative* per:

- fare avanzare lo stato dell'arte delle tecnologie di base necessarie a raccogliere, archiviare, conservare, gestire, analizzare e condividere grandi quantità di dati;
- guidare queste tecnologie per dare slancio alla ricerca nelle scienze e nell'ingegneria, rafforzare la nostra sicurezza nazionale e trasformare l'insegnamento e l'apprendimento;
- espandere la forza lavoro necessaria a sviluppare e sfruttare le tecnologie relative ai *Big Data*".



Nel documento *Seizing Opportunities, Preserving Values* sono state identificate 5 aree strategiche di intervento:

- 1. Preservare il valore della privacy:** proteggere le informazioni personali sul mercato, sia negli USA sia attraverso policy interoperabili globali;
- 2. Educazione strutturata e responsabile:** riconoscere le scuole, specie le primarie e secondarie, come una sfera importante di utilizzo dei *Big Data* per accrescere le opportunità di apprendimento, garantendo che i dati raccolti siano utilizzati per soli fini educativi;
- 3. Big Data e discriminazione:** prevenire le nuove modalità di discriminazione che i *Big Data* e la *Big Data Analytics* potrebbero rendere praticabili, espandendo le competenze tecniche delle agenzie competenti;
- 4. Giustizia e sicurezza:** garantire un utilizzo responsabile dei *Big Data* nell'applicazione della legge e nella pubblica sicurezza;
- 5. Dati come risorsa pubblica:** riconoscere che i *Big Data* sono una risorsa anche pubblica e utilizzarla per migliorare i servizi al cittadino; investire in ricerca di base e tecnologia per favorire l'ulteriore avanzamento dei *Big Data*.



## LA PERVASIVITÀ DEI *BIG DATA*

Per comprendere la potenzialità dei *Big Data* e il loro impatto in ambiti lontani dai tradizionali settori dell'IT, tra i numerosi esempi di progetti già avviati, se ne descrivono due finanziati negli anni recenti dalla Commissione Europea: TBICARE, in ambito sanitario e VIAJEO, in ambito trasporti.

## TBICARE - un progetto *Big Data* in ambito medico

Ogni anno 1.6 milioni di cittadini europei subiscono un trauma cranico (TBI), di questi 70000 non sopravvivono e 100000 subiscono disabilità permanenti. TBI è la prima causa di disabilità permanenti nella popolazione sotto i 40 anni, e il numero di pazienti affetti è in costante aumento in Europa e nel mondo.

Iniziato nel Febbraio 2011 con un finanziamento di €3M, il progetto TBICARE si è posto come obiettivo di raccogliere i dati di centinaia di pazienti che hanno subito un trauma cranico, al fine di progettare un software che aiuti i medici nella fase di diagnosi e cura.

La scelta del trattamento corretto nelle ore immediatamente successive al trauma è cruciale, ma la diagnosi può essere difficoltosa sia per la natura complessa

del cervello sia per le peculiarità del danno subito. I ricercatori di TBICARE hanno sviluppato uno strumento in grado di combinare diversi database interattivi con modelli fisiologici relativi ai traumi cranici. In questo modo i neurologi possono confrontare il caso del loro paziente con altri simili e cercare di predire con precisione il risultato della terapia.

Il progetto è frutto del coordinamento tra il VTT Technical Research Centre in Finlandia e un consorzio che include diversi ospedali e università in Europa: GE Healthcare Ltd. (UK), Turku University Central Hospital (Finlandia), University of Cambridge (UK), Imperial College London (UK), Complexio S.a.r.L. (Francia), Kaunas University of Technology (Lituania), and GE Healthcare Finland Oy.

## VIAJEO - un progetto *Big Data* in ambito trasporti

Il progetto, cominciato nel settembre 2009 a completato nell'ottobre 2012, ha coinvolto ricercatori in Europa, Cina e Brasile. Durante il progetto è stata sviluppata e implementata una piattaforma aperta per i servizi di mobilità in grado di integrare i dati da diverse fonti, come ad esempio sensori statici degli operatori di trasporto, dati proprietari dei veicoli in movimento e informazioni sul trasporto pubblico. Per testare la versatilità degli strumenti sviluppati sono state avviate sperimentazioni in quattro città: Atene, San Paolo (Brasile), Pechino e Shanghai (Cina) ognuna con peculiari caratteristiche dei dati disponibili e necessità.

L'idea di base dietro questo progetto è che la crescente doman-

da di mobilità all'interno delle aree urbane non può essere affrontata solo con la costruzione di nuove infrastrutture a causa di mancanza di spazio, impatto ambientale e costi. Per questo motivo, negli ultimi decenni si è reso necessario trovare delle soluzioni per il trasporto pubblico che prevedano un migliore sfruttamento della rete esistente. Questo richiede la raccolta di una grande quantità di dati provenienti da diverse fonti. Uno dei punti cruciali del progetto è stato proprio raccogliere, gestire e integrare questa grande e variegata mole di dati, che sono stati resi disponibili sia agli operatori di trasporto (es. taxi) per migliorare il servizio offerto, sia ai cittadini per aiutarli a pianificare gli spostamenti.





## BIG DATA E PROSPETTIVE OCCUPAZIONALI

Sulla base di uno studio di Sas e Tech-Partnership pubblicato nel 2014 [4] ed effettuato sul mercato del lavoro nel Regno Unito il numero di offerte di lavoro per posizioni relative ai *Big Data* è cresciuto da circa 1.800 nel 2008 a circa 21.400 nel 2013, con un tasso di crescita del 41% nell'ultimo anno analizzato, contro una flessione del 9% se si considera tutto il settore IT (Fig. 1). Considerando uno scenario di crescita media, il documento citato prevede fino al 2020 una crescita dell'offerta di lavoro nei *Big Data* del 23% all'anno, contro il 19% di tutto il settore IT e il 6% globale.

La situazione in Italia è stata analizzata dal CRISP dell'Università di Milano Bicocca. Le informazioni sono estratte da una base dati che raccoglie annunci di lavoro e portali web degli attori più significativi a livello nazionale nella intermediazione telematica in materia di lavoro. Complessivamente gli annunci di lavoro sul web, per le professioni nell'area *Big Data* sono stati oltre 2 mila nel periodo che va da febbraio 2013 a marzo 2016.

Il trend complessivo per l'area *Big Data* è in aumento; in particolar modo nel I trimestre 2016 si osserva un aumento del 97% rispetto allo stesso periodo dell'anno 2014. Rispetto all'anno 2015 l'aumento è del 73%.

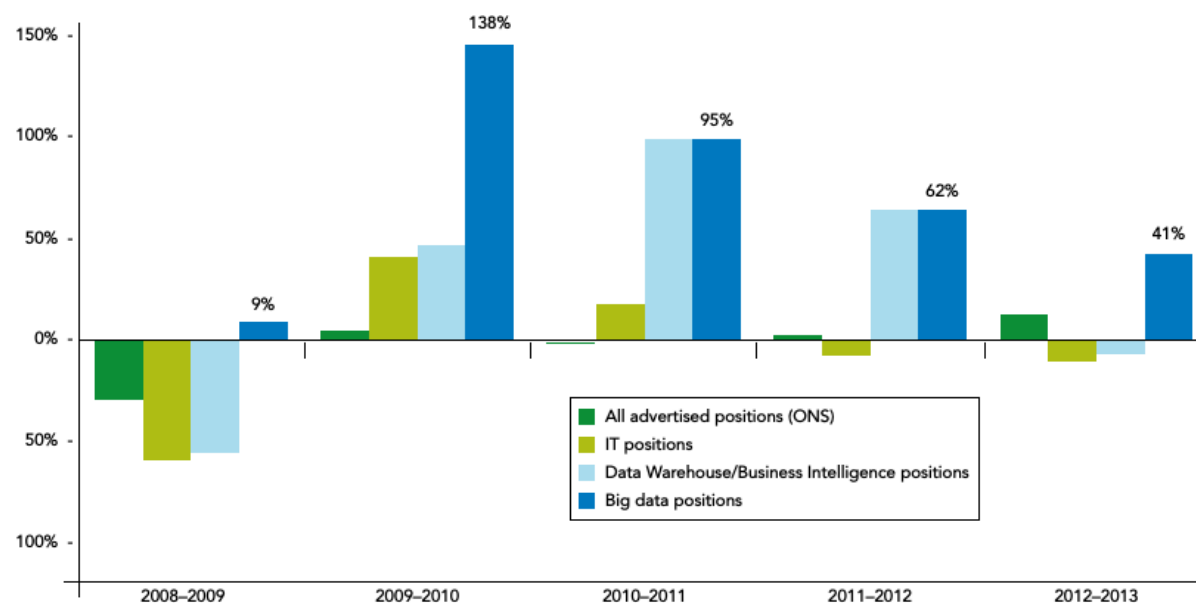
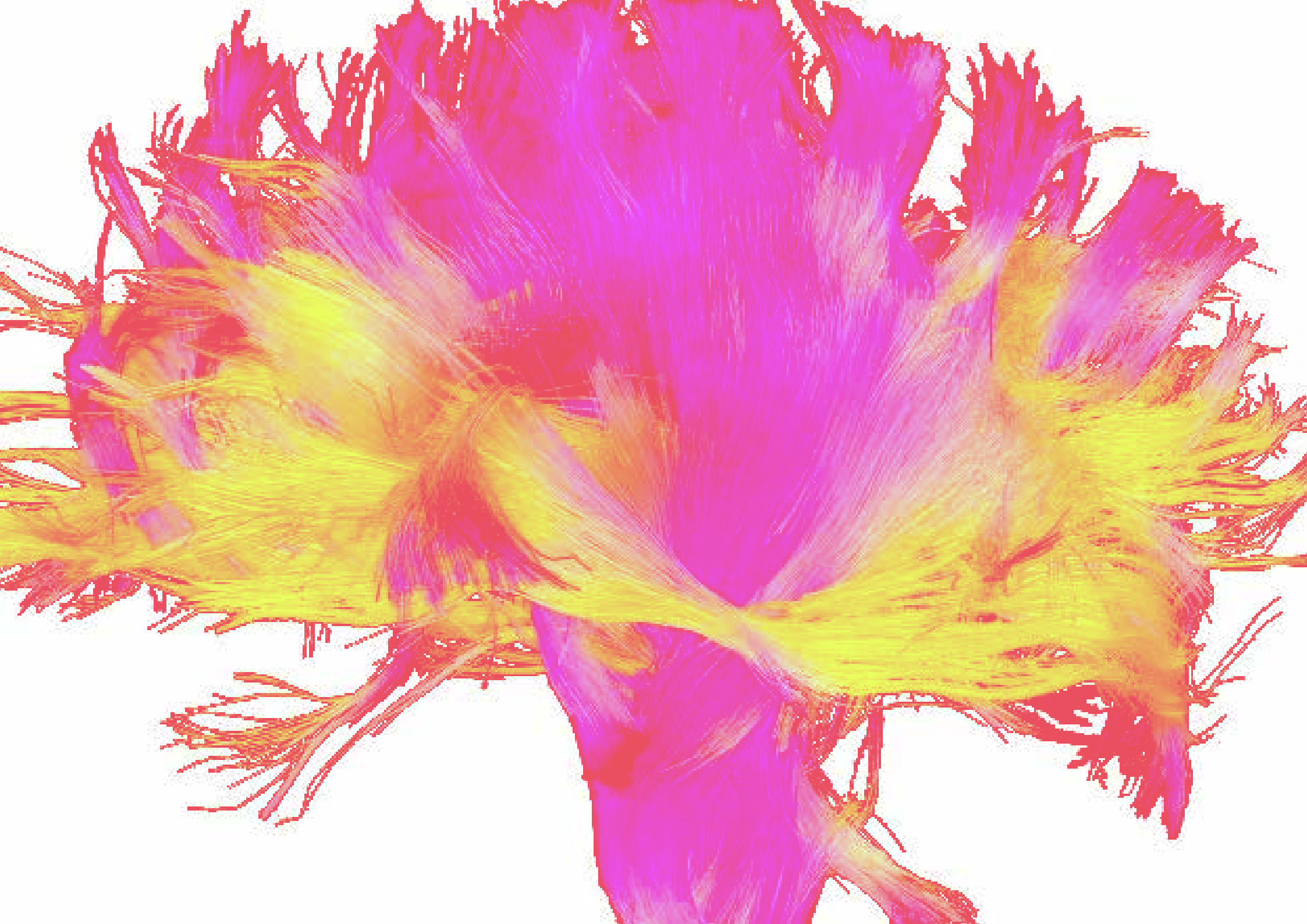


Figura 1: tasso di crescita annuale nella domanda di lavoro nel 2013.  
Fonte: Tech Partnership - Sas [4]





**2 —**

**LA FORMA —**

**ZIONE E I**

**BIG DATA**



# 1 Cosa succede in Italia e negli altri Paesi

La relativa novità dei *Big Data* si riflette a livello internazionale in un'offerta formativa ancora poco matura e conseguentemente poco omogenea. Per lo stesso motivo, non si hanno a disposizione dati statistici sui laureati e sugli sbocchi professionali. Sono stati in particolare esaminati i paesi con le economie più avanzate, in quanto ambiente naturale di questo paradigma tecnologico.

## I PAESI EUROPEI

**In Europa**, il Regno Unito è di gran lunga il paese più attivo nel campo con corsi di laurea presenti nelle più importanti università, come Oxford, Warwick, Cambridge, UCL, Edinburgh.

L'Università di Warwick, in particolare, vanta un Bachelor in Data Science e due Master (MSc) direttamente collegati ai *Big Data*, oltre a diversi master (MSc) in argomenti affini. L'università di Edinburgo ha avviato nel 2014 un percorso dottorale in Data Science

(MRes+PhD) della durata di 4 anni, e prevede di formare 10 studenti all'anno per 5 anni.

L'University College di Londra (UCL) ha un master (MSc) in Data Science con l'obiettivo di formare figure di Data Analyst, *Big Data* Architect, SQL Developer, Business Analyst.

Degno di nota il Master of Arts (MA) *Big Data in culture and society* del King's College di Londra, attivo dal 2016, che offre un anno di formazione in *Big Data* specifico per i beni culturali e le scienze sociali.

Per quanto riguarda l'Europa continentale, la Germania è quella col tessuto più vitale, con diversi percorsi in fase di attivazione. Da menzionare la Jacobs University Bremen con un master biennale (MSc) e la Technische Universität München con due master biennali (MSc) uno in Data Engineering e l'altro in Data Analytics. Meno significativa l'offerta didattica in Francia, anche se sono presenti vari corsi professionalizzanti post-laurea e alcuni Massive Online Open Courses - MOOC (Telecom ParisTech, École Polytechnique, Paris VI).

Vale la pena citare i consorzi EuroTech Universities e EIT. Il primo, formato da quattro importanti università tecniche europee (Danmarks Tekniske Universitet, École Polytechnique Fédérale de Lausanne, Technische Universiteit Eindhoven, Technische Universität München), ha tra le sue *focus area* la Data Science & Engineering e sta esplorando le possibili sinergie sulla formazione in tema di *Big Data*. Il secondo propone un percorso master (MSc) in data science frutto della col-

laborazione di diverse università europee: Politecnico di Milano, Technische Universiteit Eindhoven, Universidad Politecnica de Madrid, Université Nice Sophia Antipolis, Technische Universität Berlin e KTH Royal Institute of Technology.

## USA

Anche negli **Stati Uniti** l'offerta formativa strutturata in corsi di studio specializzati non è particolarmente estesa, con alcune importanti eccezioni, tra cui vale la pena citare la Michigan Data Science Initiative, che prevede lo stanziamento di 100 milioni di dollari e il reclutamento di 35 tra professori e ricercatori in 5 anni. Sono però piuttosto diffusi corsi singoli inseriti in percorsi formativi più tradizionali o generalisti e un'ampia offerta di MOOC.

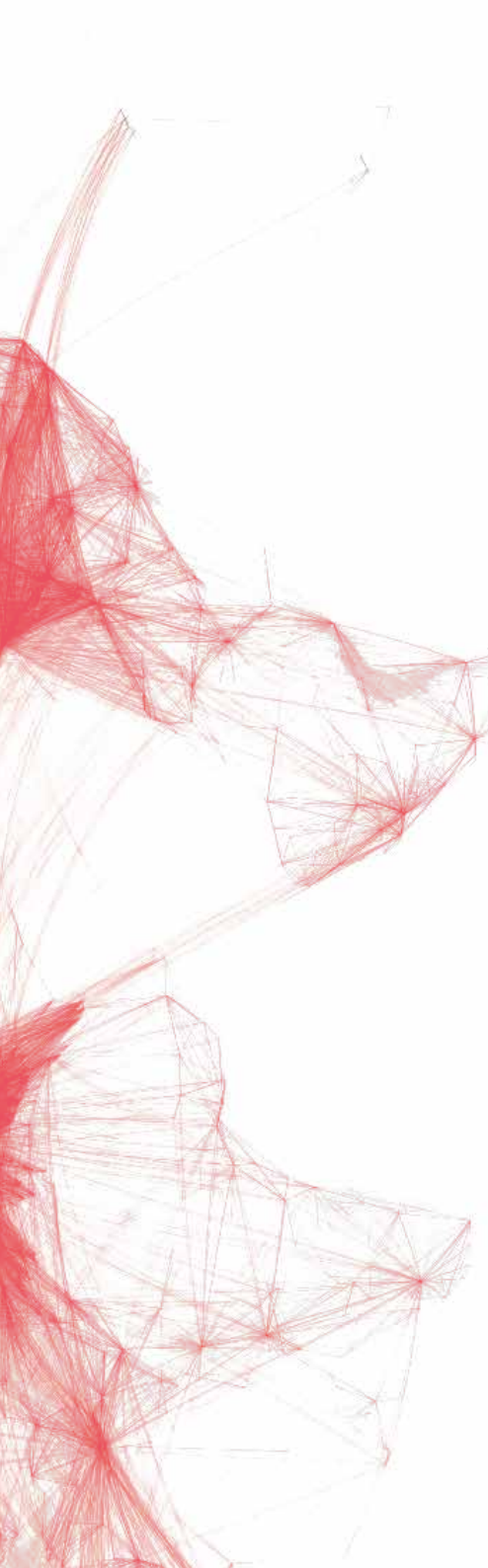
## COSA SUCCEDE IN ITALIA

Per quanto riguarda l'**Italia**, l'importanza di formare la figura del *data scientist* è stata recepita a diversi livelli da parte delle università. Il gruppo di lavoro ha in primo luogo censito l'offerta esistente attraverso i database del MIUR, ma ha anche richiesto autosegnalazioni per le iniziative in corso di definizione, intervistato alcuni dei referenti di questi corsi e ottenuto materiale informativo da parte degli stessi.

Prevedibilmente, l'offerta didattica è nata in grande parte all'interno dei dipartimenti di informatica e ingegneria informatica, i quali molto frequentemente propongono uno o più corsi tematici su argomenti più o meno strettamente legati ai *Big Data*. Tali corsi sono disponibili per gli studenti impegnati in percorsi di studio "tradizionali", per lo più magistrali, ma anche per lauree di primo livello e dottorati. Non è invece diffusa la presenza di tali corsi all'interno di corsi di laurea di aree diverse.

Dal punto di vista dei corsi di laurea appositamente pensati per la formazione del *data scientist*, ne sono stati censiti undici. L'intrinseca multidisciplinarietà degli argomenti e la varietà degli obiettivi scelti nei diversi atenei si riflettono in frequenti collaborazioni tra diversi dipartimenti e in una certa varietà nella strutturazione dell'offerta formativa.





**Dei corsi censiti, uno solo è una laurea triennale, tre sono lauree magistrali, e i restanti sono organizzati come curricula all'interno di corsi di laurea di area informatica o ingegneristica. Il resto dell'offerta è costituito da master sia di primo sia di secondo livello piuttosto disomogenei per ricchezza dell'offerta e livello. Non uniforme anche la presenza di un'attività e di un'infrastruttura di ricerca specifica per il settore dei *Big Data* e della loro analisi.**

In tutti i corsi considerati si registra la partecipazione di studenti con *background* molto eterogenei. Significativamente, è stata evidenziata la difficoltà di partecipare con successo a questi percorsi da parte di studenti con formazione umanistica: ostacolo principale è la mancanza delle conoscenze di base richieste come prerequisiti alla partecipazione, ma anche l'orientamento dei corsi attualmente disponibili, per lo più focalizzati sull'ambito economico. In generale è stata evidenziata la difficoltà di accogliere studenti a causa della rigidità dei requisiti all'ammissione ai corsi di studio magistrali e ai master: anche studenti provenienti da ambiti scientifici (ad es. medicina) possiedono spesso competenze matematiche molto lontane da quelle conseguite durante un corso triennale di ambito informatico e sono quindi esclusi dai requisiti di ammissione alle lauree magistrali.

Un altro fattore che è stato indicato come un freno alla formazione in questo ambito è la mancanza delle in-

frastrutture informatiche necessarie a trattare quantità di dati paragonabili ai casi reali, nonché, in alcuni casi, la difficile reperibilità di dati di sufficiente qualità e dimensione.

Nelle tabelle seguenti è riassunta l'offerta, con la distinzione tra atenei che offrono corsi di laurea o master di I e II livello dedicati, inclusi i programmi non ancora avviati, e quelli che offrono solo corsi singoli all'interno di programmi di laurea triennale (LT), magistrale (LM) o dottorato (D).

Tabella 1 Offerta di/in corsi di laurea e master in Italia				
Università	LT	LM	Master	Offerta in altri corsi di laurea
Bologna			I	LT, LM
Firenze			I, II	LT, LM
Genova	●	● <sup>1</sup>		D
L'Aquila		● <sup>1</sup>		
Milano - Bicocca			I	
Milano - Bocconi		● <sup>1</sup>		
Milano - Politecnico		● <sup>1</sup>		
Molise			I	
Padova		● <sup>2</sup>		
Pisa		●	II	
Roma - Luiss			I	
Roma - Sapienza		●	II	LT, LM
Roma - Tor Vergata		● <sup>1</sup>	II	
Siena		● <sup>1 2</sup>		
Torino		●	II	
Torino - Politecnico			II <sup>3</sup>	LM
Trieste - SISSA			II	
Venezia	● <sup>1</sup>	● <sup>1</sup>		

Tabella 2 Offerta di insegnamenti all'interno di corsi di laurea o dottorato			
Bolzano	LM	Padova	LT, LM
Cassino	LM	Palermo	LM
Ferrara	LT, LM, D	Pavia	LM
Insubria	D	Politecnico delle Marche	LM
Lucca - IMT	D	Roma TRE	LM
Messina	LT, LM	Salerno	LM
Milano - Statale	LM, D	Trento	LM
Napoli - Federico II	LM		

Legenda: <sup>1</sup> curriculum, <sup>2</sup> futuro, <sup>3</sup> non avviato nell'anno in corso

## ALCUNI ESEMPI

### **Statistica matematica e trattamento informatico dei dati**

Università: Genova - Dipartimento: Matematica  
Titolo: Laurea Triennale  
Classe: L-35  
Primo anno di corso: 2001  
Lingua: Italiano  
[www.dima.unige.it/SMID](http://www.dima.unige.it/SMID)

L'obiettivo del corso di laurea è formare figure professionali con competenze statistiche e solide conoscenze matematiche e informatiche.

La formazione di base è quindi di carattere matematico e informatico con approfondimenti specifici in ambito probabilistico e statistico. Viene dato ampio spazio alle applicazioni della statistica a problemi di carattere biologico, fisico, medico, ambientale ed economico. Il percorso universitario include uno stage in azienda che permette allo studente di affrontare esperienze tipiche del mondo del lavoro.

Il corso di studi offre una formazione professionale che permette l'inserimento nel mondo del lavoro già al conseguimento della laurea. La prosecuzione degli studi in lauree magistrali, tipicamente LM delle classi in discipline statistiche, è uno sbocco che riguarda più della metà dei laureati SMID.

### **Business informatics**

Università: Pisa  
Dipartimento: Informatica  
Altri dipartimenti coinvolti: Economia e Management, Giurisprudenza, Istituto di Scienze e Tecnologie dell'Informazione del CNR  
Titolo: Laurea Magistrale  
Classe: LM-18  
Primo anno di corso: 2002  
Lingua: Inglese  
[www.di.unipi.it/it/didattica/wbi-lm](http://www.di.unipi.it/it/didattica/wbi-lm)

Pensato come corso interdisciplinare fra informatica ed economia, nonostante l'enfasi sugli aspetti economici, il CdS presta attenzione anche agli aspetti sociologici e socio-politici relative all'uso dei *Big Data*, oltre che a quelli più decisamente tecnologici e statistico-matematici. Il CdS è aperto a studenti di qualsiasi provenienza, anche umanistica, ed è stato valutata come la migliore laurea magistrale di Pisa in termini di placement dei laureati.

## Data Science

Università: Sapienza, Roma

Dipartimento: Scienze Statistiche

Altri dipartimenti coinvolti: Informatica, Ingegneria dell'Informazione, Elettronica e Telecomunicazioni, Ingegneria Informatica, Automatica e Gestionale "Antonio Ruberti"

Titolo: Laurea Magistrale

Classe: LM-91

Primo anno di corso: 2015

Lingua: Inglese

[datascience.i3s.uniroma1.it/](http://datascience.i3s.uniroma1.it/)

Il CdS è aperto a studenti provenienti da Matematica, Fisica, Economia, Informatica e Statistica e prevede un colloquio di ammissione. Il programma del corso ha un'impronta forse maggiormente multidisciplinare rispetto al CdS organizzato dall'Università di Torino, con più spazio dedicato alla organizzazione e gestione dei dati oltre che all'analisi. L'organizzazione del CdS trae ispirazione dal corso analogo in Data Science organizzato presso la New York University, uno dei primi ad essere creati in una grande Università nordamericana. Anche questo CdS è organizzato in collaborazione con diversi partner industriali, inclusi alcuni vendor di tecnologie di storage, come EMC2.

## Stochastics and Data Science

Università: Torino

Dipartimento: Matematica

Altri dipartimenti coinvolti: Scienze Economiche-Sociali e Matematico-Statistiche, Informatica

Titolo: Laurea magistrale

Classe: LM-40

Primo anno di corso: 2015

Lingua: Inglese

[www.master-sds.unito.it/](http://www.master-sds.unito.it/)

L'accesso al corso è regolato dalla verifica del curriculum ed eventualmente da un colloquio volto a determinare preparazione e motivazione dei candidati. Il CdS prevede corsi in tre ambiti principali (Matematica, Statistica e Informatica) e punta in particolare sull'aspetto matematico e teorico nella formazione dei data scientist. Tra le collaborazioni con partner industriali spicca quella molto stretta con Intesa-San Paolo. È inoltre in cantiere la progettazione di un programma di Dottorato sulle stesse tematiche.

## QoLexity - Measuring, Monitoring and Analysis of Quality of Life and its Complexity

Università: Firenze

Dipartimento: Statistica, Informatica, Applicazioni, "G. Parenti"

Altri dipartimenti coinvolti: ISTAT

Titolo: Master II livello

Primo anno di corso: 2013

Lingua: Inglese

<http://www.disia.unifi.it/vp-51-master-ii-livello-html>

L'obiettivo del master, frutto della collaborazione tra l'università di Firenze e l'ISTAT, è di definire, misurare, monitorare e analizzare la qualità della vita in termini quantitativi, coinvolgendo diverse discipline quali filosofia, sociologia, psicologia, statistica, economia e scienze politiche. Il master mira a formare professionisti con competenze statistiche applicabili a livello analitico, manageriale e di comunicazione.





## **Business Intelligence e Big Data Analytics**

Università: Milano - Bicocca  
Dipartimento: CRISP  
Titolo: Master I livello  
Primo anno di corso: 2010  
Lingua: Inglese  
<http://www.crisp-org.it/bimaster/>

Di indirizzo più statistico/economico, ha come target persone che hanno completato il loro percorso di studi ma che trovano necessario/interessante acquisire competenze in *Data Science*. Unico tra quelli censiti, dà spazio, tra le altre competenze, alla moderna disciplina del *data journalism*, argomento sul quale gli organizzatori del master hanno proposto una *summer school*. Il master è comunque rivolto a studenti di statistica, informatica, ingegneria o economia.

## **Gestione dei Dati e dei Documenti Digitali nelle Amministrazioni Pubbliche e Private**

Università: Molise  
Dipartimento: Bioscienze e Territorio  
Titolo: Master I livello  
Primo anno di corso: 2016  
Lingua: Italiano

<http://dipbioter.unimol.it/didattica/altri-corsi-di-studio/master/master-in-big-data/>

L'obiettivo del master è quello di creare figure professionali in grado di gestire basi di dati strutturate e non, trattare i dati sensibili, utilizzare sistemi cloud-based, sfruttare tecniche avanzate di statistica e di analisi dati, essere in grado di fornire indicazioni strategiche ai management aziendali. Il master affianca a lezioni frontali segmenti di apprendimento con piattaforme online.

## **SoBigData - European Research Infrastructure on Social Mining & Big Data Ecosystem**

Università: Consorzio Europeo coordinato dal CNR di Pisa  
[www.sobigdata.eu](http://www.sobigdata.eu)

La Commissione Europea, nel programma Horizon 2020, ha scelto il consorzio a guida italiana per la nuova infrastruttura di ricerca sui *Big Data*. SoBigData è una rete di centri europei che mette a disposizione dati, strumenti e competenze di data scientist per condurre grandi esperimenti di *Big Data* analytics da parte di ricercatori, innovatori, startupper, policy-makers, istituzioni pubbliche ed imprese private. Esperimenti rivolti ad estrarre senso dalle tracce digitali delle attività umane registrate appunto nei *Big Data*, ed usare la conoscenza estratta per nuove scoperte scientifiche, ad esempio in campo sociale ed economico, per nuovi servizi e prodotti, o a supporto delle decisioni. SoBigData ha l'ambizione di diventare un ecosistema di tecnologie e di persone in grado di liberare il potenziale dei *Big Data* come bene pubblico, aperto e accessibile, in un quadro etico di valori imprescindibili: trasparenza dei processi e delle finalità, rispetto della privacy, fiducia e responsabilità. Fra i compiti centrali dell'infrastruttura c'è quello di fornire una piattaforma per la formazione dottorale nella data science, in cui i dottorandi potranno sperimentare nuovi metodi analitici e nuove applicazioni in tutti gli ambiti disciplinari

## **Big Data @ Polito**

Univeristà: Torino - Politecnico

<http://bigdata.polito.it/>

Il laboratorio BigData@polito è il frutto della collaborazione di quattro dipartimenti dell'ateneo: Elettronica e Telecomunicazioni (DET), Automatica e Informatica (DAUIN), Scienze Matematiche (DISMA), ingegneria Gestionale e della Produzione (DIGEP) e offre sia risorse computazionali a ricercatori interessati al portare avanti progetti di ricerca nel campo *Big Data*, sia le competenze e l'esperienza maturata all'interno dei dipartimenti coinvolti. Ogni dipartimento è specializzato ad un approccio alla tematica affine al contesto in cui opera. In particolare, all'interno del DET, il gruppo TNG (Telecommunication Networks Group) ha utilizzato le tecnologie *Big Data* per analizzare grandi quantità di dati raccolti da Internet. Il DBDMG (DataBase and Data Mining Group), nel DAUIN, si è focalizzato nelle tecniche di data mining e gestione di database, con particolare attenzione ai limiti degli algoritmi tradizionali che scaturiscono quando si ha a che fare con *Big Data*.

Il focus del DISMA è a livello metodologico e di metodi matematici, ruolo che svolge anche fornendo attività di consulenza a società esterne.

Infine, il DIGEP ha focalizzato la sua ricerca nell'identificare quale sia l'impatto dei Big Data nel mondo industriale, con particolare attenzione alle capacità da sviluppare nelle divisioni ICT, alle opportunità di creazione di valore accessibili alle aziende applicando il paradigma Big Data, e, a livello macroscopico, all'evoluzione delle dinamiche di competizione e di creazione di valore dovuta all'avvento dell'era dei dati.

## **CRISP**

Università: Milano - Bicocca

<http://www.crisp-org.it/>

I ricercatori afferenti al CRISP hanno competenze multidisciplinari (statistica, informatica, economia) con l'obiettivo comune di estrarre conoscenza (e quindi valore) dai dati per supportare il processo decisionale. La principale attività di ricerca del CRISP si focalizza quindi su tutti i passi del processo di estrazione di informazione dai dati: dalla data selection, alla pre-elaborazione fino ai processi di classificazione, mining e visualizzazione, con una grande attenzione alla ricerca applicata in progetti regionali, nazionali ed europei.

Dal 2015 il CRISP è alla guida di un tender Europeo per l'analisi di Big Data del mercato del lavoro europeo, commissionato dal Cedefop volto alla creazione di un prototipo per il collezionamento, la classificazione, l'analisi e delle professioni e skill delle Web job vacancy pubblicate nei principali siti dei 5 paesi partecipanti (Italia, Germania, Repubblica Ceca, UK, Irlanda). L'obiettivo è di supportare la labour market intelligence dell'ente (circa 4,5 milioni di job vacancy collezionate negli ultimi 4 mesi).

Nel 2015 il CRISP ha promosso la Summer School on BI&Big Data analytics cui hanno partecipato di più di 40 tra studenti di dottorato, dipendenti di aziende e lecturer che sono intervenuti su temi come Information Extraction, Data Integration, Graph DB, *Big Data* e SNA.

Il CRISP è membro attivo del *Big Data* Group dell'European Network on Regional Labour Market Monitoring.





## 2. Imparare i Big Data a scuola

La pervasività e la rilevanza del fenomeno dei *Big Data* richiedono un intervento articolato su più livelli. Non sarebbe infatti sufficiente un'azione tesa alla creazione di figure di specialisti –pur necessarie–, perché questo non inciderebbe nelle diverse aree disciplinari e professionali dove mancherebbe la capacità di **capire l'importanza dei Big Data** nel settore specifico e quindi la volontà di utilizzare queste professionalità aggiuntive. Al tempo stesso l'obiettivo qui sostenuto di favorire la formazione di veri *digital natives* con una specifica sensibilità al valore del dato in un'ottica *Big Data* avrebbe tempi troppo lunghi per mettere il paese in condizioni di competere nei tempi necessari con questi nuovi strumenti per la modernizzazione e lo sviluppo: da qui l'esigenza di prevedere anche percorsi formativi adatti per i professionisti nei diversi settori per integrarne le competenze con questo nuovo insieme di approcci.

Nel seguito quindi verranno proposte azioni specifiche a diversi livelli partendo dalla Scuola secondaria e verranno evidenziate le possibili implicazioni sulla modalità stessa di apprendimento e sulle opportunità di valutazione sia dei discenti che dei docenti. Per quanto concerne la Scuola, in particolare, verranno richiamati alcuni elementi del Piano Nazionale Scuola Digitale

(di seguito PNSD), il documento strategico con cui il MIUR ha definito il posizionamento del sistema scolastico nell'era digitale in quanto già delineano azioni in linea con gli esiti dell'indagine qui svolta.

### **2.1 I BIG DATA A SCUOLA: TRA COMPETENZE DEGLI STUDENTI E AMBIENTI DI APPRENDIMENTO**

Nel processo di alfabetizzazione digitale della scuola, in particolare di quella secondaria, i nuovi traguardi formativi devono includere i *Big Data* come strumento abilitante per lo sviluppo di nuove competenze trasversali. Questa previsione si inserisce nel contesto del PNSD nel quale si afferma che la sfida formativa "è oggi relativa in primo luogo alla capacità di reperire, comprendere, descrivere, utilizzare, produrre informazione complessa e strutturata, tanto nell'ambito scientifico e tecnologico quanto *in quello umanistico e sociale [...] la produzione di contenuti digitali diventa sempre più articolata e complessa, e richiede competenze adeguate: competenze logiche e computazionali, competenze tecnologiche e operative, competenze argomentative, semantiche e interpretative*".

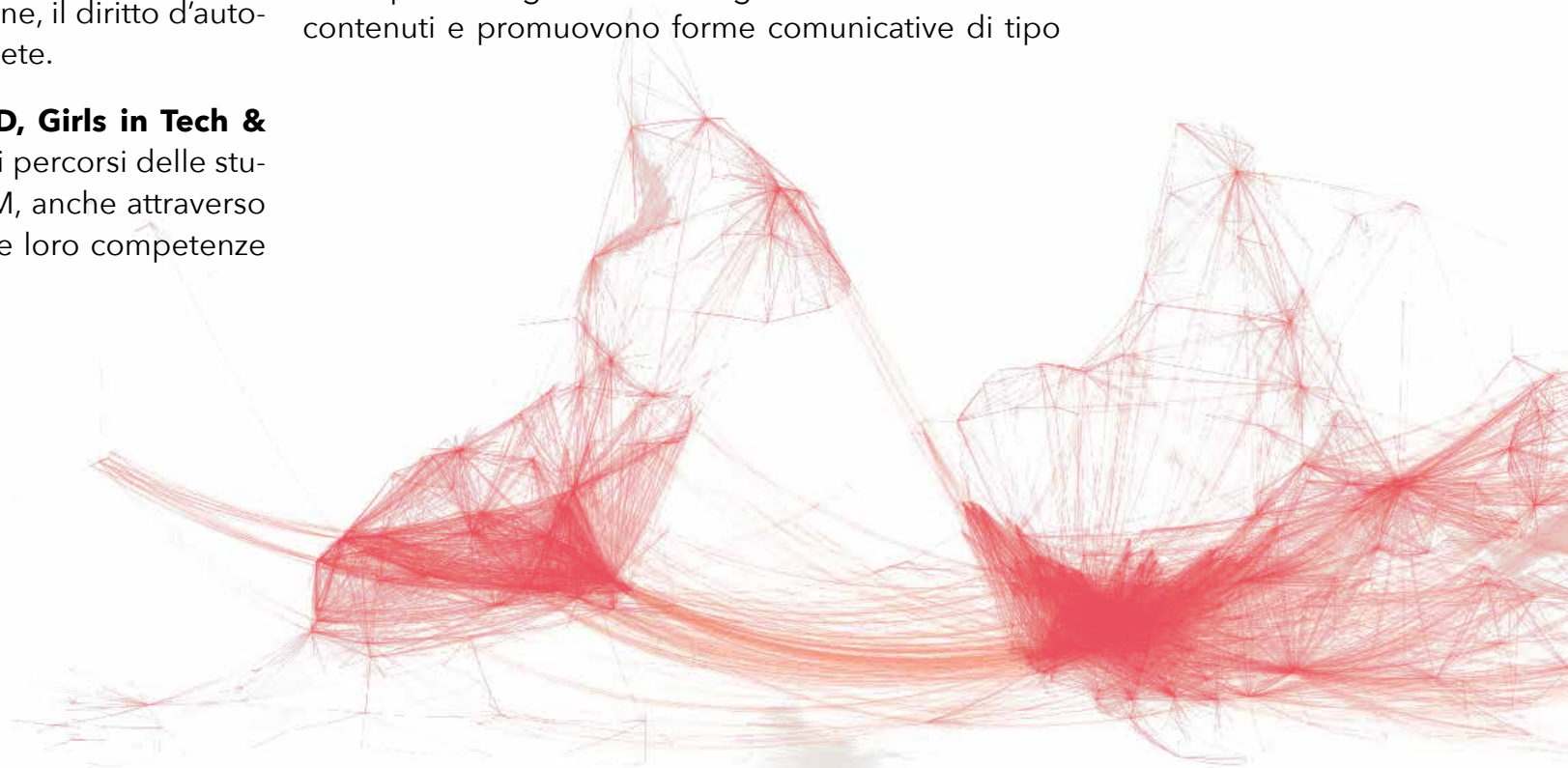
I *nativi digitali*, come peraltro raccomandato anche dall'OCSE, grazie a questi percorsi formativi devono **trasformarsi da meri consumatori in "consumatori critici" e "produttori" di contenuti e architetture digitali**, in grado di sviluppare competenze trasversali a ogni settore e ambito occupazionale; in grado di risolvere problemi, concretizzare le idee, acquisire auto-


*mia di giudizio, pensiero creativo, consapevolezza delle proprie capacità, duttilità e flessibilità nella ricerca di soluzioni.*

Tornando al PNSD, l'azione "Scenari Innovativi" finanzia 25 curricula didattici applicabili a tutta la scuola, con due specifici percorsi dedicati ai dati (*Big e Open Data*), dando particolare attenzione all'educazione alla comprensione e gestione delle dinamiche di generazione, analisi, rappresentazione e riuso dei dati. Tra gli altri "scenari didattici" finanziati e prodotti da questa linea d'azione, figurano inoltre competenze essenziali e propedeutiche ai *Big Data* e al lavoro in ambienti digitali, come il pensiero computazionale, *l'information literacy* (le caratteristiche e la filosofia dell'informazione, l'organizzazione dell'informazione, il diritto d'autore), *l'Internet of Things* e i diritti in Rete.

Attraverso l'**Azione #20 del PNSD, Girls in Tech & Science**, il MIUR punta a orientare i percorsi delle studentesse verso le discipline STEAM, anche attraverso azioni per rafforzare e accelerare le loro competenze in queste materie.

Oltre ai *Big Data* oggetto e strumento didattico, vi è poi una dimensione di particolare interesse legata a un drastico **rinnovamento delle metodologie formative**, alla diffusione degli **ambienti digitali**, ossia di sistemi software per la produzione di contenuti didattici e l'organizzazione di percorsi formativi *web based finalizzati ad agevolare e semplificare i processi di apprendimento*. All'interno di tali ambienti virtuali, le piattaforme di apprendimento o *Learning Management System (LMS)* possono promuovere una concezione 'circolare' della comunicazione tra docente e discente, e potenziare i servizi digitali scuola-famiglia-studente. Questi strumenti non solo incentivano una pratica sistematica delle abilità digitali di base, ma soprattutto generano una gestione interattiva dei contenuti e promuovono forme comunicative di tipo





cooperativo e negoziale e nuove opportunità per la verifica in tempo reale dell'efficacia degli strumenti formativi e del progresso di singoli e gruppi di discenti. Tali sistemi, infatti, favorendo la dematerializzazione del materiale formativo, portano a costituire *repository* centralizzati dei dati e documenti generati sia dalle interazioni scuola-famiglia, sia dai processi di esercitazione e valutazione (prove, esercitazioni, dati sui risultati di apprendimento ma anche materiali più prettamente burocratico-amministrativi) **che possono agevolare un successivo riuso dei materiali a fini valutativi, di monitoraggio e ricerca didattica** [12]. I dati relativi alle interazioni, anche solo i tracciati di interfacciamento dei discenti con la piattaforma, possono essere resi immediatamente disponibili all'analisi per **fornire un quadro complessivo dei comportamenti della classe e ottimizzare la strategia di apprendimento quasi in tempo reale** [13]. In tal senso, applicazioni di *Learning Analytics* sui *Big Data* raccolti dalle piattaforme, possono rappresentare soluzioni specializzate per valutare la qualità della proposta didattica proprio sulla base della risposta cognitiva dei discenti e proporre innovazione all'interno degli obiettivi di apprendimento predefiniti dai curricula nazionali [14].

In questa stessa prospettiva è possibile raffinare *learning analytics* secondo modalità "*adaptive learning technology*", selezionare e ricostruire i dati secondo gli impulsi generati dal singolo soggetto per lo sviluppo di corsi personalizzati che registrino in anticipo possibili carenze e migliorino le performance nella preparazione degli studenti [15]. Al tempo stesso, ottimizzando lo sviluppo delle interazioni di tali piattaforme con il web, ovvero con *social networks* o altri media, i nuovi paradigmi tecnologici permettono di sperimentare l'impiego di *sentiment analysis* per 'ascoltare', capire le percezioni degli utenti, quali elementi di stimolo per raffinare il servizio erogato.

## 2.2 | Big Data nell'Alta formazione: laurea, laurea magistrale, PhD, master

La formazione universitaria rappresenta il livello ideale per fornire gli strumenti richiesti dal fenomeno emergente dei *Big Data*. L'analisi presentata nella sezione 2.1 mostra che gli atenei, anche italiani, hanno iniziato ad elaborare un'offerta strutturata in questa direzione con l'obiettivo di formare i *data scientist*: nel seguito di questa sezione saranno esplicitate alcune proposte sulle caratteristiche ottimali di questi percorsi.

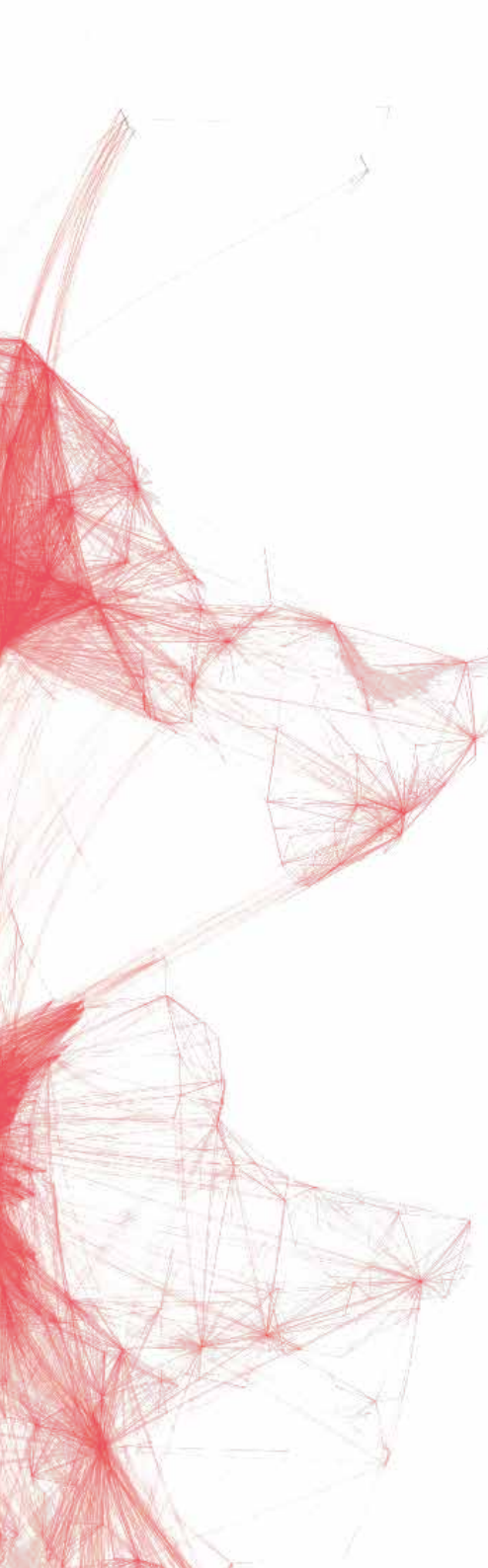
**Preliminarmente, dunque, appare necessario inserire moduli di *data science* in tutti i corsi di studio, a partire dalle lauree triennali, per esporre gli studenti al cambio di paradigma che investe tutte le discipline spingendo verso un approccio scientifico *data driven*.**

In tal modo, la modellizzazione *top-down* dei fenomeni trova una nuova sintesi con la scoperta di conoscenze *bottom-up*, che emergono dalle grandi masse di dati disponibili - esperimenti sui dati visti non solo come validazione di teorie e modelli, ma anche come **scoperta** di pattern emergenti che suggeriscono agli scienziati nuove teorie e nuovi modelli, in grado di spiegare più a fondo la complessità dei fenomeni sociali, economici, biologici, tecnologici, culturali.

L'obiettivo è espandere la capacità degli studenti di comprendere l'importanza dei dati in generale e nel loro percorso di studi in particolare, ovvero:

- i meccanismi che governano la natura dei dati strutturati e non strutturati, la loro produzione, acquisizione e integrazione da fonti eterogenee, la loro gestione e cura;
- le modalità con cui si estrae conoscenza dai dati, non solo con l'analisi statistica tradizionale, ma anche con metodi di *data mining* e *machine learning* che consentono di apprendere dai dati (e validare) modelli predittivi e pattern emergenti;
- le modalità narrative e visuali con cui comunicare le conoscenze;
- i temi etici legati all'uso dei dati e delle conoscenze, i modelli organizzativi per il loro sfruttamento, e come tutti questo possa influenzare profondamente la vita degli studenti, professionale e non.





Questi obiettivi formativi richiederanno la riserva di crediti formativi all'interno dei programmi didattici dei percorsi di laurea esistenti che ovviamente dovranno essere ricavati all'interno dei limiti dell'offerta complessiva. In molti casi lo spazio all'interno delle lauree triennali potrà essere -almeno in parte- ottenuto rimodulando le attività formative per le competenze informatiche e digitali di base, ormai patrimonio comune delle generazioni degli attuali studenti, specie in un contesto quale quello qui delineato che include già la scuola secondaria.

Una riflessione sui paradigmi formativi ottimali nei diversi settori deve essere promossa anche attraverso l'incentivazione di progetti comuni tra *data scientist* ed esperti dei diversi settori (sanità, economia, scienze sociali, scienze umane...).

Elemento di attenzione è il reperimento delle necessarie figure di docenza. Il background del docente di *data science* è all'incrocio fra informatica, fisica, statistica e matematica. Occorrono competenze profonde di *data management* e costruzione di processi computazionali, unite all'analisi e modellizzazione matematica e statistica di fenomeni complessi. Queste figure docenti oggi provengono principalmente dall'ambito informatico (*data mining e machine learning*), fisico (sistemi complessi) e statistico (*statistical learning*), e si sono formate sul campo attraverso preziose contaminazioni disciplinari.

**Uno specifico programma di reclutamento da parte del Ministero appare auspicabile per sopperire alla prevedibile scarsità di personale docente al momento disponibile nei ranghi degli atenei.**

In alcuni casi potrebbero essere utilmente creati centri di ateneo capaci di fornire i crediti di ambito *Big Data* previsti nei diversi indirizzi.

Una situazione diversa è identificabile per i percorsi di **laurea magistrale**. Infatti, se sicuramente ulteriori elementi di data science sono necessari all'interno dell'offerta di tutti i percorsi, qui, e non a livello di laurea triennale, si colloca il livello ideale per la formazione di specialisti in *data science*. Un aspetto centrale della programmazione di questi percorsi di laurea è la necessità di renderli adatti ad accogliere laureati provenienti da background molto diversi per formare degli **specialisti di settore** capaci di utilizzare in maniera avanzata e indipendente i nuovi strumenti del mondo dei *Big Data* **in specifici settori scientifici e professionali**.

Conseguentemente tali percorsi di **laurea magistrale** potrebbero in molti casi essere inseriti in una varietà di classi disciplinari distinte a seconda dello specifico settore nel quale il data scientist svolgerà la sua attività (ad es. settore biomedico, scienze sociali o scienze umane).

Anche se classi interdisciplinari appaiono di interesse, al fine di non creare ostacoli in ingresso, né barriere disciplinari per il prosieguo dell'eventuale vita accade-

mica degli studenti, una maggiore varietà di contesti disciplinari è desiderabile.

Tali percorsi richiedono la presenza di competenze piuttosto diversificate e di una numerosità adeguata di personale docente con specifiche attività di ricerca di settore per cui adeguate condizioni quali-quantitative dovranno essere poste per l'accREDITAMENTO di queste nuove lauree magistrali.

Considerazioni simili, ma rafforzate si applicano alla possibilità di attivare percorsi di **formazione dottorale**. Qui la presenza di una documentata capacità di ricerca dell'ateneo ospite è condizione essenziale e nuovi dottorati potrebbero gemmare da esperienze di successo in dottorati esistenti in fisica o informatica attraverso il coinvolgimento di esperienze di alcuni settori di applicazione delle nuove metodologie. Uno specifico **piano di finanziamento di borse dottorali** per data scientist potrebbe avviare in maniera decisa queste nuove esperienze. Oltre a un limitato numero di questi dottorati focalizzati sulla figura del *data scientist*, potrebbero essere utilmente previste borse riservate a candidati con un *background* specifico da inserire all'interno di percorsi più convenzionali di diversi settori (economia, farmacologia, etc.).

Numerosi **master** sono già offerti, come sempre in questa tipologia di corsi, con caratteristiche piuttosto disomogenee. All'interno di un disegno formativo coerente sui *Big Data*, tuttavia, il ruolo dei master può essere di grande interesse. La finalità principale di questi percorsi professionalizzanti dovrebbe essere la formazione continua, favorendo la collaborazione con le imprese. Nell'attesa che giunga a compimento la formazione dei *big-data native*, infatti, percorsi che integrino le competenze dei professionisti già oggi attivi in determinati settori con i metodi della data science sono di grande interesse. Non è quindi raccomandato il formato del master sui *Big Data* tout court, ma piuttosto una varietà di percorsi intensivi sull'utilizzo delle nuove metodologie riservato agli esperti di un determinato settore disciplinare/professionale. Esempi evidenti sono master aperti a economisti, scienziati politici, farmacologi, epidemiologi, etc. Al momento i master sono poco strutturati nel sistema formativo italiano e il MIUR dovrebbe considerare un intervento mirato sull'**accREDITAMENTO** di questi percorsi per garantirne la qualità e corretta strutturazione.





# 4 Modifiche al sistema formativo, valutativo, di monitoraggio

L'interazione sistematica con i supporti informatici determina una dimensione digitale generalizzata delle informazioni ministeriali tanto di quelle elaborate dai percorsi didattici (formativi e valutativi), quanto di quelli indirettamente prodotti dal *management* amministrativo, economico-finanziario, che diviene il presupposto "strutturale" per lo sviluppo di capacità analitiche evolute di *Big Data*, sia in itinere che *ex post*, dei dati e delle informazioni raccolte.

Questo processo tecnicamente si incardina sulla capacità del sistema di potenziare un ambiente di *data* e *content management* che coniughi condivisione, co-creazione e aggregazione dell'informazione, dove la fase di produzione del dato e quelle di esplorazione, osservazione, riflessione, interpretazione possano essere fortemente connesse.

Il raccogliere i singoli *database* istituzionali in un unico sistema *knowledge based* può rappresentare un'opportunità per migliorare la capacità di analisi delle informazioni e la loro interpretazione. I nuovi modelli di

*management* permettono di integrare più facilmente i dati attraverso scambi *machine-to-machine*, riutilizzare l'informazione raccolta in contesti di analisi diversi da quelli specifici di origine e generare, sulla base dei dati aggregati, un più ampio spettro informativo. Il fine è riuscire a **centralizzare i dati istituzionali in un unico sistema, arricchirli avvalendosi anche di informazioni recuperate da altre articolazioni ministeriali, dal comparto pubblico in genere oppure da contesti esterni, affinché l'informazione elaborata da processi di Content Analysis possa essere reinterpretata e ridistribuita a favore della collettività.**

Sotto questo profilo il paradigma *Big Data* sembra offrire nuove frontiere di indagine all'interno delle azioni di sistema promosse dal Ministero nei comparti della Istruzione, Università e Ricerca per lo studio dei fenomeni di interesse, l'analisi dei servizi erogati e l'efficienza dei processi.

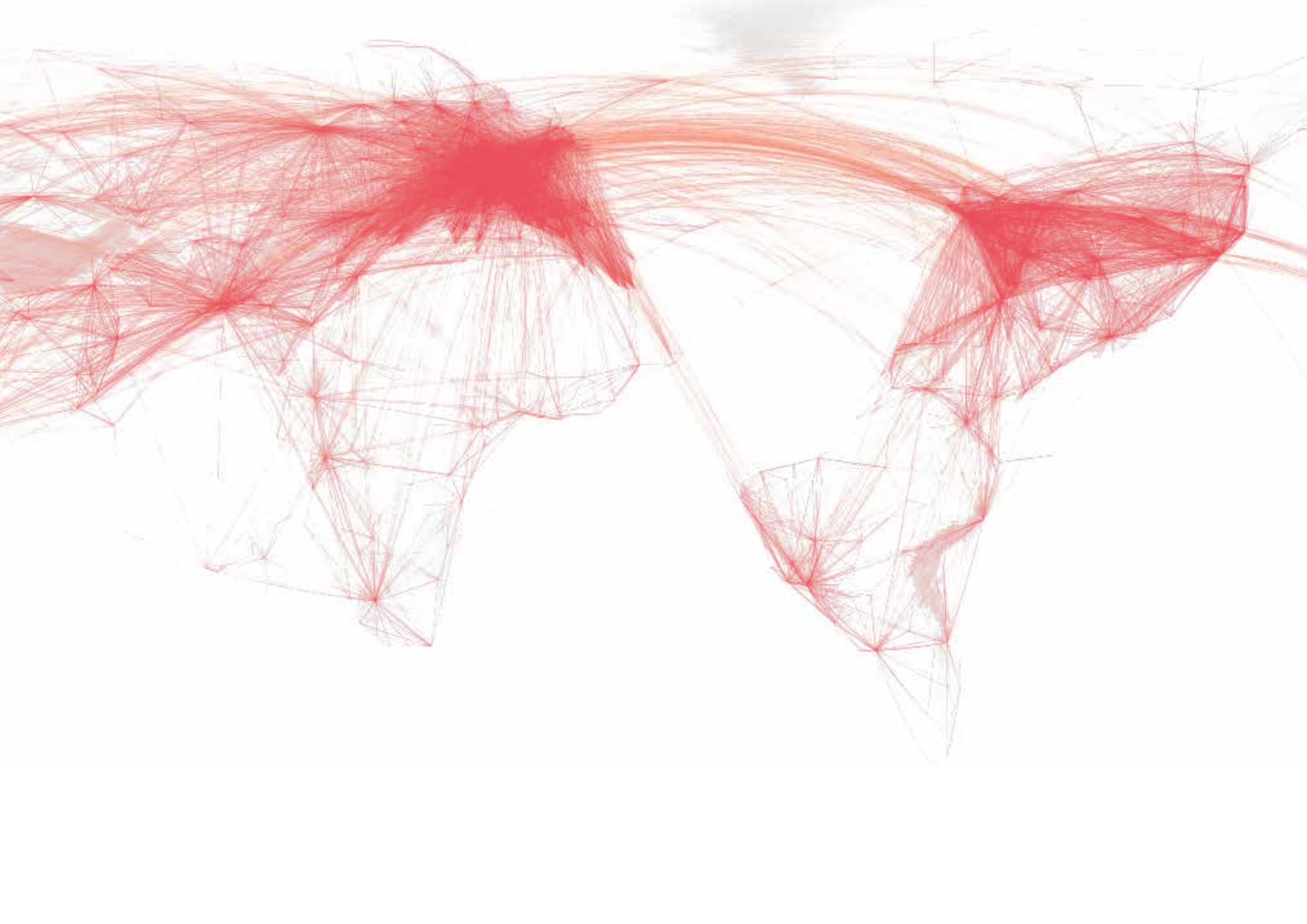
Le iniziative a supporto della trasformazione del sistema nel suo complesso oppure di alcune articolazioni interne, promosse soprattutto per incentivare i cambiamenti nella strutturazione dei servizi, in termini di miglioramento e potenziamento organizzativi, possono oggi essere arricchite con nuove metodologie di analisi. Vale a dire che certi interventi istituzionali possono essere programmati e orchestrati anche alla luce di analisi di settore multidimensionali condotte in parallelo, anche attraverso i diversi *data set* di comparto, prendendo in considerazione contesti di indagine che altrimenti rimarrebbero inesplorati.

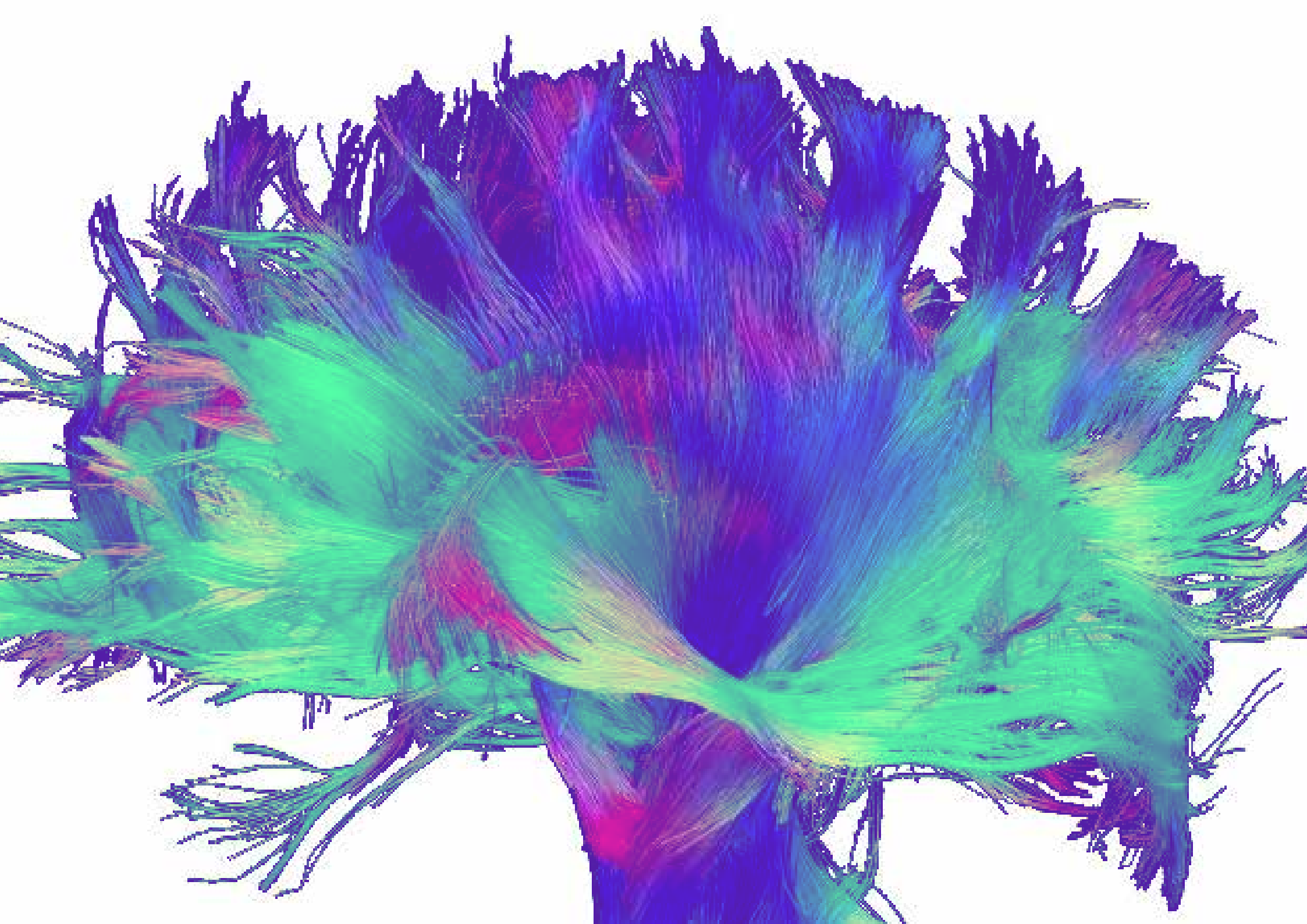
Le analisi di settore proposte dalle nuove tecnologie, grazie all'eterogeneità delle fonti elaborate, permettono di ricostruire la struttura reticolare di un determinato fenomeno nel suo aspetto più complesso e completo, e possono risultare particolarmente strategiche per ripensare alcuni processi di indagine e renderli maggiormente funzionali ad osservare, comprendere e regolare le dinamiche di cambiamento e di innovazione in atto nel sistema nazionale.

In tal senso, **gli stessi processi di monitoraggio ministeriali oppure quelli di valutazione e autovalutazione** promossi dal sistema scolastico, attraverso l'impiego di strategie di trattamento automatico di *Big Data*, **possono essere ripensati. Si possono cioè trasformare da meri strumenti tecnico-amministrativi, in più articolate procedure di analisi** che si avvalgono di informazioni tempestive, provenienti da contesti più ampi di quelli tradizionali. Offrire un'indagine secondo una prospettiva più articolata, multidimensionale, permette di studiare un aspetto del settore educativo per individuare quali possano essere i fattori che lo determinano o maggiormente lo condizionano e ricostruire una più ampia visione di insieme anche delle molteplici dimensioni socio-economiche e culturali a cui il fenomeno studiato può essere connesso. Tutto ciò può infine suggerire nuove interpretazioni dei dati, rendere più tempestivi i meccanismi di controllo delle performance istituzionali e più calibrate le proposte di intervento.

Le nuove strategie di analisi risultano particolarmente congeniali anche per promuovere l'analisi sistematica di un fenomeno per guidarne le riforme istituzionali e il loro consolidamento nei diversi contesti sociali (cd. *data driven policy making*). In particolare, offrendo analisi di sistema su alcuni aspetti del fenomeno educativo, incrociando quanto promosso dal decisore (*processo top down*) con quanto percepito e vissuto dagli attori del sistema (*processo bottom up*), è possibile tentare di prevenire alcuni limiti del sistema indotti da quadri normativi di contesto che spesso soffrono di un certo divario tra le linee programmatiche e le esigenze e i limiti dell'agito.

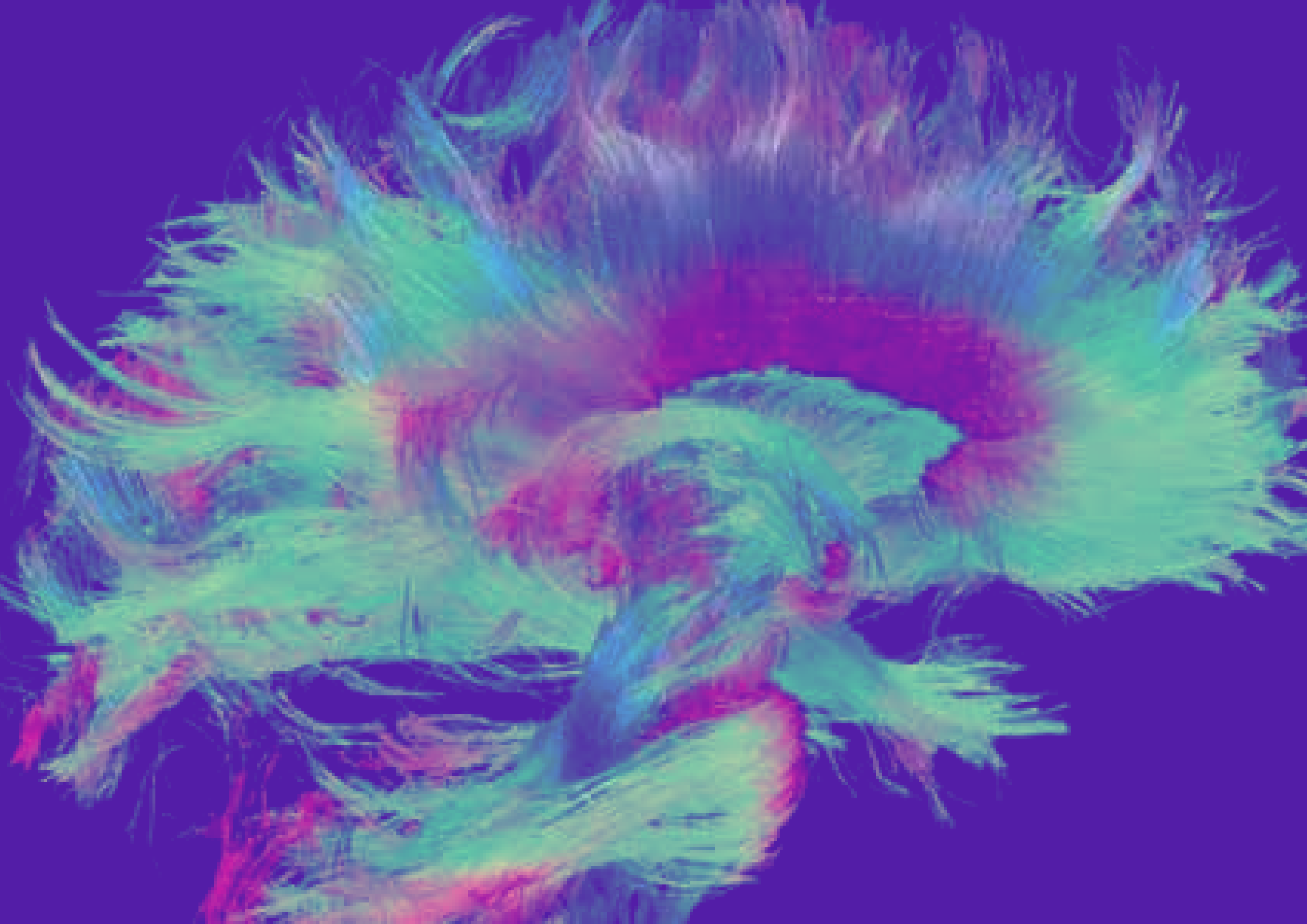






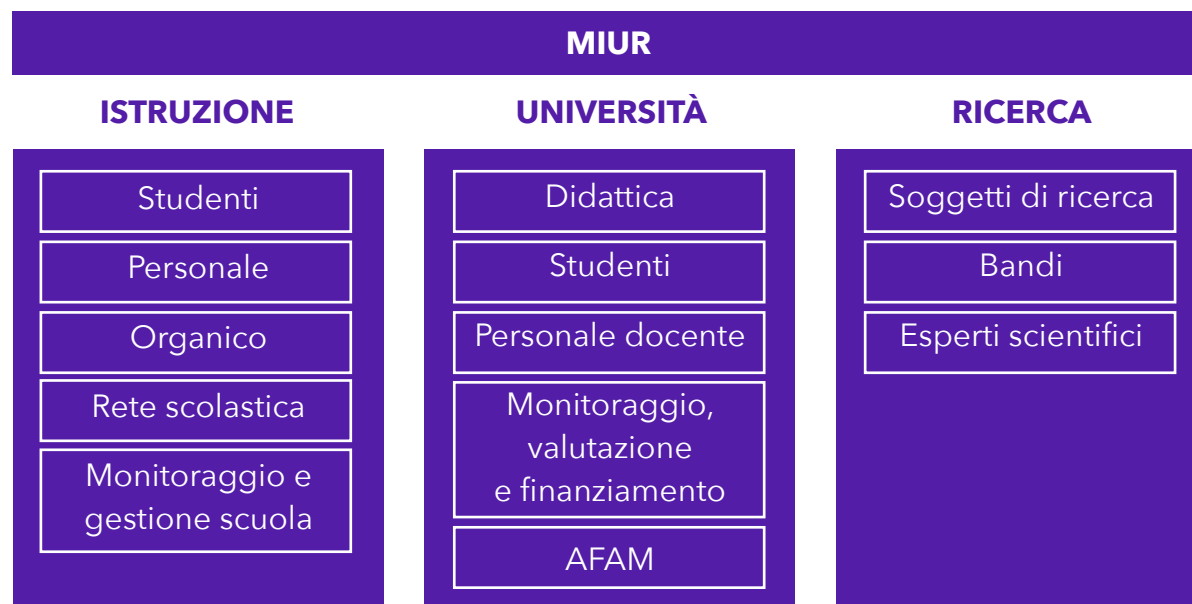
**3 —**

**IL MIUR E  
IL PATRI—  
MONITORO DEI  
BIG DATA**



# 1 Le basi di dati MIUR: il censimento

Il sistema informativo del Ministero dell'Istruzione, Università e Ricerca è organizzato in tre sottosistemi: **Istruzione, Università e Ricerca** che saranno analizzati separatamente.



## Sistema Informativo MIUR Istruzione - Aree e Processi

Il sistema informativo dell'**ISTRUZIONE** comprende tre macro-aree: **Alunni/Studenti, Scuole e Personale**, su ciascuna delle quali insistono processi e dati connessi.

In generale, la macro-area **Alunni/studenti** è descritta dai dati relativi a: dati anagrafici, frequenze-abbandoni-trasferimenti, percorsi didattici, alternanza scuola-lavoro, scrutini, esami di Stato, denunce infortuni.

Macro Area	I Livello	II Livello	Responsabilità del dato
Studenti	Anagrafe alunni scuole primarie	Anagrafe	DG CASIS
		Frequenze e Scrutini	DG CASIS
		Infortuni Inail	DG CASIS
	Anagrafe studenti scuole secondarie di I grado	Anagrafe	DG CASIS
		Frequenze e Scrutini	DG CASIS
		Esami di Stato (III media)	DG CASIS
		Infortuni Inail	DG CASIS
	Anagrafe studenti scuole secondarie di II grado	Anagrafe	DG CASIS
		Frequenze e Scrutini	DG CASIS
		Esami di Stato (Maturità)	DG CASIS
		Infortuni Inail	DG CASIS

Riguardo alle **Scuole**, le aree interessate sono le Istituzioni Scolastiche statali, i Plessi scolastici (statali e non statali), l'Edilizia Scolastica, la Gestione Finanziaria e il Sistema di Autovalutazione.

In particolare, i dati che afferiscono a tali aree riguardano rispettivamente:

- Per le Istituzioni scolastiche: anagrafica, dirigenza (intesa come sede del Dirigente scolastico), Osservatorio Tecnologico, Fatturazione Elettronica, Fondi Europei, denunce infortuni, Organico.
- Per i Plessi, statali e non statali: anagrafica, tipologia di plesso, quadri orario.
- Per l'Edilizia scolastica: anagrafica, dati tecnici (certificazioni, strutture, locali, etc.).
- Per la Gestione Finanziaria: risorse in entrata, risorse in uscita, tipologia di finanziamento, Piano dei Conti della Spesa.
- Per il Sistema di Autovalutazione: anagrafica, dati del contesto economico-sociale, esiti (risultati scolastici nella scuola e nel successivo percorso di studi/nel mondo del lavoro), processi (attività svolte dalla scuola, organizzazione della scuola ed interazioni con il contesto di riferimento).



Macro Area	I Livello	II Livello	Responsabilità del dato
Scuole	Istituzioni scolastiche statali	Anagrafica	DG CASIS
		Dirigenza	DG PER
		Osservatorio tecnologico	DG EFID
		Fatturazione Elettronica	DG CASIS
		Fondi Europei	DG EFID
		Infortuni Inail	DG CASIS
		Pianta Organica	Dip Istruzione
Scuole	Plessi (Statali e non statali)	Anagrafica	DG CASIS
		Tipologia Plesso	DG CASIS
		Libri di Testo	DG OSV
		Quadri Orario	DG OSV/CASIS
	Edilizia	Anagrafica	DG EFID
		Certificazioni	DG EFID
		Locali	DG EFID

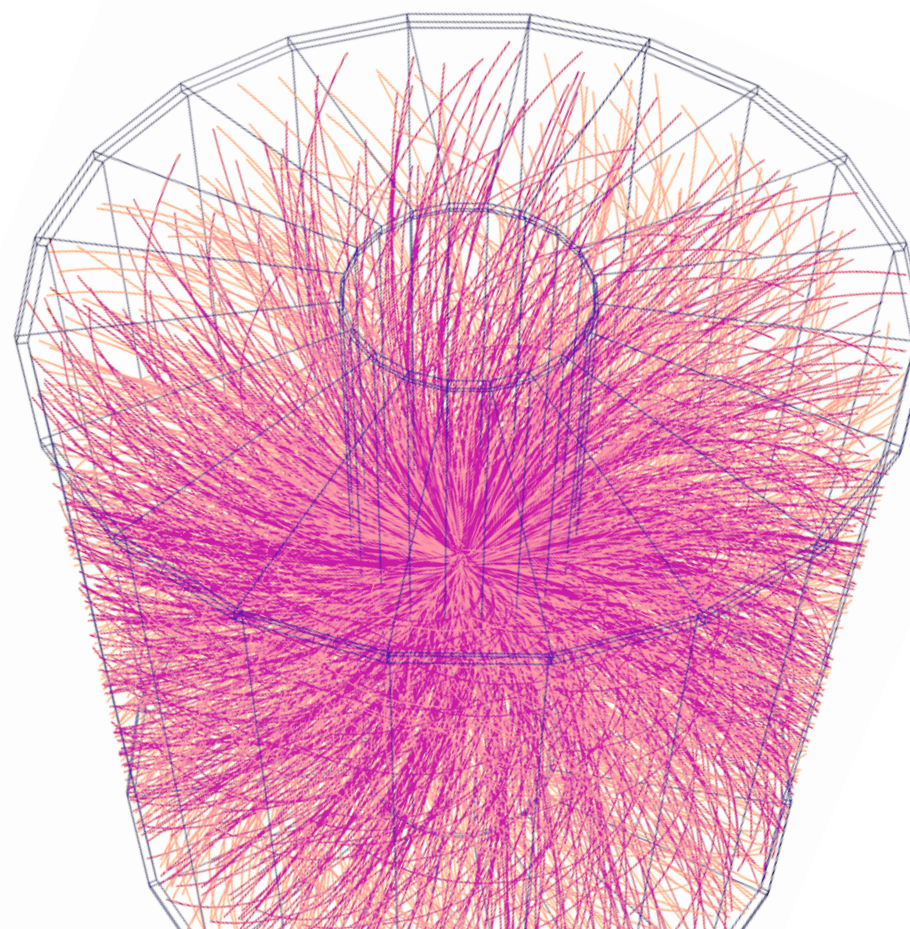
Macro Area	I Livello	II Livello	Responsabilità del dato
Scuole	Gestione finanziaria	Anagrafica	DG CASIS
		Risorse in entrata	DG RUF
		Risorse in uscita	DG RUF
		Tipologia di finanziamento	DG RUF
		Piano dei conti della spesa	DG RUF
	Sistema di autovalutazione	Anagrafica	DG CASIS
		Contesto	DG CASIS
		Esiti	DG CASIS
		Processi	DG CASIS

Nella macro-area del **Personale** si distinguono i seguenti ambiti e i relativi dati che vi afferiscono:

- Incarico, servizio, denunce infortuni.
- **Personale Scuola di ruolo** (include Docenti, Personale Amministrativo, Tecnico e Ausiliare (ATA), Personale Educativo, Insegnanti di Religione Cattolica): dati anagrafici, dati del ruolo, titolarità e servizio, contratti, retribuzioni, assenze, professionalità docente, denunce infortuni.
- **Personale Scuola non di ruolo** (include Docenti, Personale ATA, Personale Educativo): dati anagrafici, retribuzioni, dati di servizio, contratti, graduatorie, assenze, denunce infortuni.
- **Personale dell'Amministrazione Centrale e Periferica**: dati anagrafici, retribuzioni, collocazione, assenze, revisori contabili, tipologia contratto, denunce infortuni.

Macro Area	I Livello	II Livello	Responsabilità del dato
<b>Personale Istruzione</b>	Dirigenti Scolastici	Anagrafica	DG PER
		Retribuzioni	DG RUF
		Incarico	DG PER
		Servizio	DG PER
		Infortuni Inail	DG CASIS
<b>Personale Istruzione</b>	Personale scuola di ruolo (Docenti, ATA, PED, IRC)	Anagrafica	DG PER
		Ruolo	DG PER
		Titolarità	DG PER
		Servizio	DG PER
		Contratti	DG PER
		Retribuzioni	DG RUF
		Assenze	DG PER
		Professionalità docente	DG PER
		Infortuni Inail	DG CASIS

Macro Area	I Livello	II Livello	Responsabilità del dato
Personale Istruzione	Personale scuola non di ruolo (Docenti, ATA, PED)	Anagrafica	DG PER
		Retribuzioni	DG RUF
		Servizio	DG PER
		Contratti	DG PER
		Graduatorie	DG PER
		Assenze	DG PER
		Infortuni Inail	DG CASIS
Personale Istruzione	Personale Amministrativo	Anagrafica	DG RUF
		Retribuzioni	DG RUF
		Collocazione	DG RUF
		Assenze	DG RUF
		Revisori Contabili	DG RUF
		Tipologia Contratto	DG RUF
		Infortuni Inail	DG CASIS

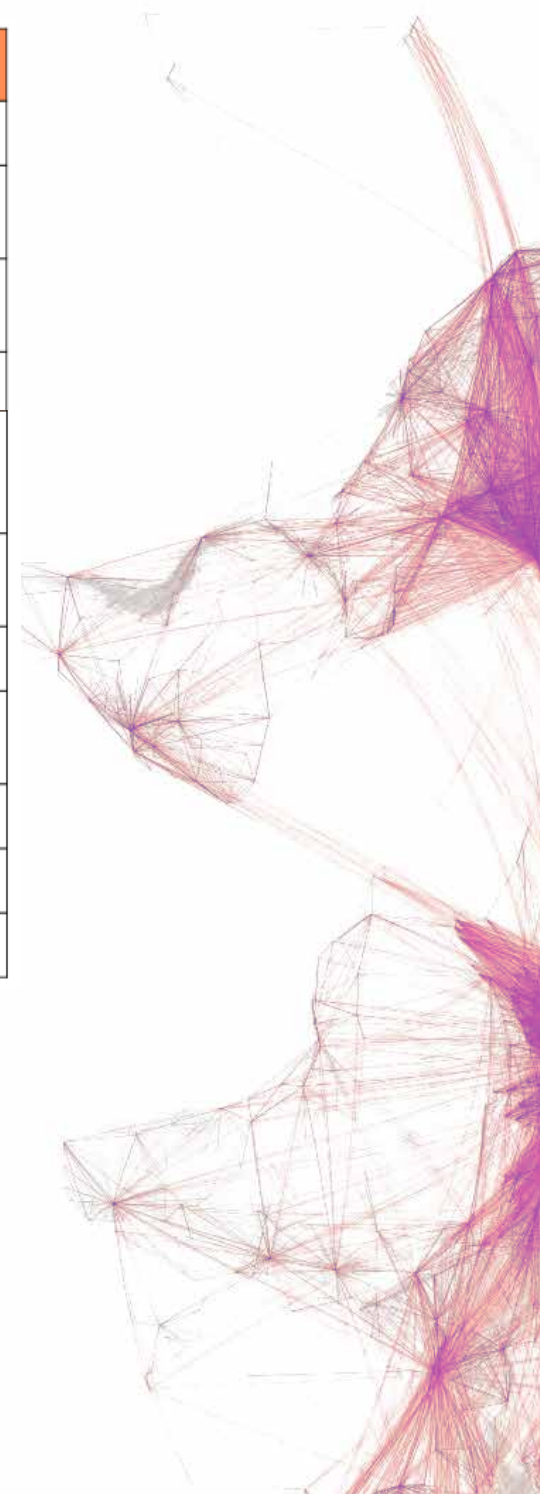


## Sistema Informativo MIUR Università e Ricerca - Aree e Processi

Il sistema informativo dell'**UNIVERSITÀ** è classificabile in tre macro-aree: Studenti, Didattica- Finanziamenti-Strutture e Personale, su ciascuna delle quali insistono processi e dati connessi.

In generale, la macro-area **Studenti** è descritta dai dati relativi a: dati anagrafici, carriera, gettito della contribuzione studentesca, interventi a favore degli studenti erogati dalle università e dagli Enti per il diritto allo studio. In particolare, solo per i corsi di Laurea, dati relativi alla mobilità (Erasmus); limitatamente ai corsi di Laurea ed ai TFA (Tirocinio Formativo Attivo), dati relativi alle prove di ammissione.

Macro Area	I Livello	II Livello	Responsabilità del dato
Studenti	Lauree, Lauree Magistrali e Ciclo unico	Accesso programmato	DGSINFS
		Anagrafe nazionale studenti	DG CASIS/ DGSINFS/ DGFIS
		Rilevazione Istruzione Universitaria	DG CASIS
		Erasmus	DGSINFS /INDIRE
Studenti	Post Laurea	Iscritti e diplomati Master I e II Liv (rilev. statistica)	DG CASIS
		Iscritti e diplomati Scuole specializ (rilev. statistica)	DG CASIS
		SSPWEB - Studenti Iscritti	DGSINFS
		Iscritti e dottorati Dottorati di Ricerca	DG CASIS
	AFAM	Iscritti e diplomati AFAM	DG CASIS
	ITS	Banca Dati ITS	DGOSV/INDIRE
	TFA	Iscritti TFA	DGSINFS



Riguardo alla macro-area **Didattica, Finanziamenti e Strutture**, le aree interessate sono: Università, Alta Formazione Artistica e Musicale (AFAM), Offerta Formativa, Valutazione-Programmazione-Monitoraggio.

In particolare, i dati che afferiscono a tali aree riguardano rispettivamente:

- Per l'Università: strutture delle Università e le loro principali caratteristiche, dati finalizzati alla ripartizione del Fondo Integrativo alle Regioni per il Diritto allo Studio; dati delle strutture, delle spese sostenute, del piano finanziario e del programma delle attività sportive degli studenti universitari ai fini del calcolo del contributo da erogare (CUSWEB); collegi universitari legalmente riconosciuti.
- Per gli AFAM: strutture e loro principali caratteristiche.
- Per l'Offerta Formativa: offerta formativa dei Corsi di Laurea accreditati (SUA), offerta formativa dei Dottorati di Ricerca accreditati (OFF\_Dott), piano di dettaglio delle attività formative (SSPWEB), informazioni relative a modalità, contenuti ed esiti anonimi delle prove di ammissione alle scuole di specializzazione in Medicina, a livello nazionale.
- Per la Valutazione, Programmazione e Monitoraggio: dati di bilancio degli atenei (Conti consuntivi), costo del personale delle Università statali (DALIA), dati per il calcolo del Fondo del Finanziamento Ordinario (FFO), dati necessari per le attività di

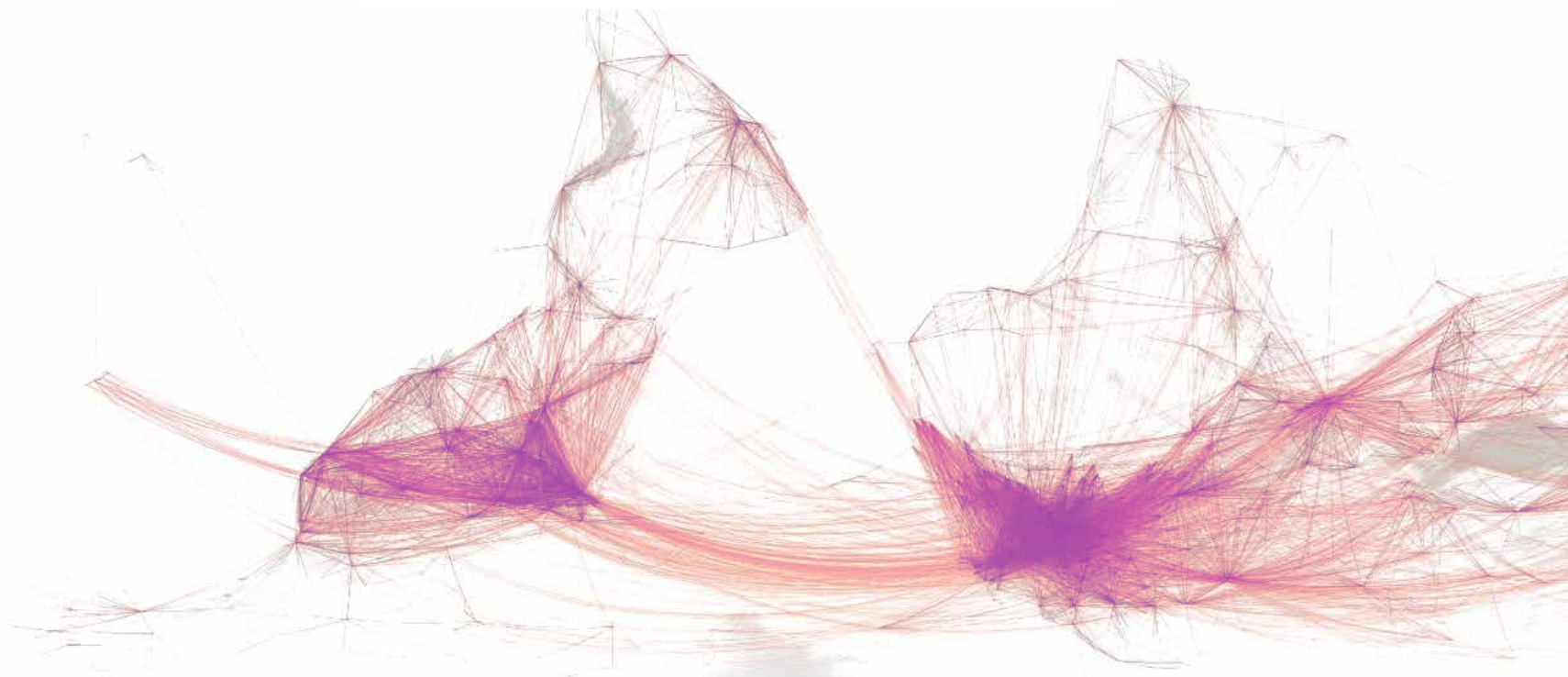
coordinamento, attuazione, monitoraggio e valutazione della Programmazione Triennale (PRO3).

Macro Area	I Livello	II Livello	Responsabilità del dato	
Didattica, Finanziamenti e strutture	Università	Strutture	DGFIS	
		Fondo integrativo alle regioni	DGSINFS	
		Diritto allo studio	DG CASIS	
		CUSWEB	DGSINFS	
		Contribuzione ed interventi Atenei	DG CASIS	
		Collegi Universitari	DGSINFS	
AFAM	AFAM	Strutture AFAM	DGSINFS	
		Contribuzione ed interventi AFAM	DG CASIS	
Didattica, Finanziamenti e strutture	Offerta formativa	Offerta formativa SUA (Rad, Requisiti minimi)	DGSINFS	
	Offerta formativa Post Laurea	SSPWEB - Attività Formative, Tirocini, Convenzioni e Sedi	DGSINFS	
		Scuole di specializzazione in medicina	DGSINFS	
		OFF_Dott (off formativa Dottorati di Ricerca)	DGSINFS	
	Valutazione, Programmazione e Monitoraggio	Valutazione, Programmazione e Monitoraggio	Conti consuntivi	DGFIS
			DALIA	DGFIS
			FFO	DGFIS
PRO3			DGFIS	

Nella macro-area del **Personale** si distinguono i seguenti ambiti ed i relativi dati che vi afferiscono:

- **AFAM**: anagrafica dei docenti di ruolo presso le istituzioni statali dell'AFAM.
- **Università**: anagrafica dei docenti di ruolo, domande di Abilitazione Scientifica Nazionale (ASN), titolari di assegni di ricerca, docenti delle Scuole di Specializzazione in Psicoterapia (SSPWEB - Docenti); personale docente a contratto e personale T.A delle università e degli istituti universitari statali e non statali, programmazione delle assunzioni nelle Università Statali (PROPER).

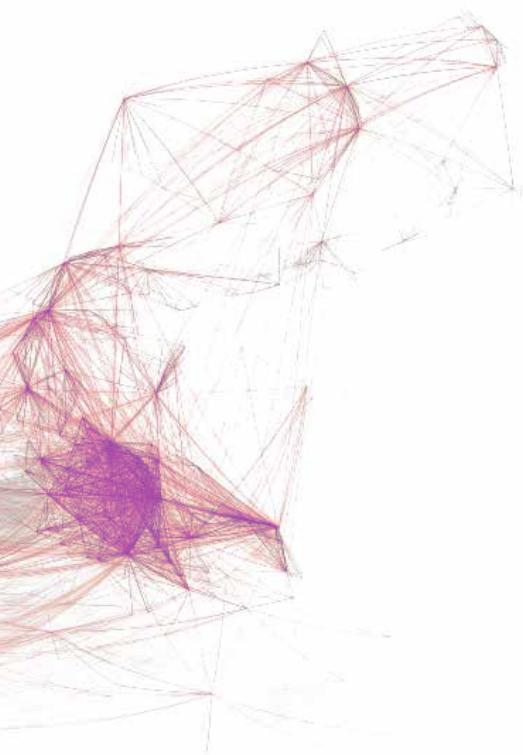
Macro Area	I Livello	II Livello	Responsabilità del dato
Personale Università	Università	DB Docenti di ruolo (Professori e Ricercatori) Università	DG CASIS
		Docenti a contratto e personale T.A.	DG CASIS
		PROPER	DGFIS
		ASN - Abilitazione Scientifica Nazionale	DGFIS
		SSPWEB - Docenti	DGSINFIS
		Assegnisti	DGFIS
	AFAM	AFAM - rilevazione	DG CASIS
		DB Docenti di ruolo afam	DGFIS



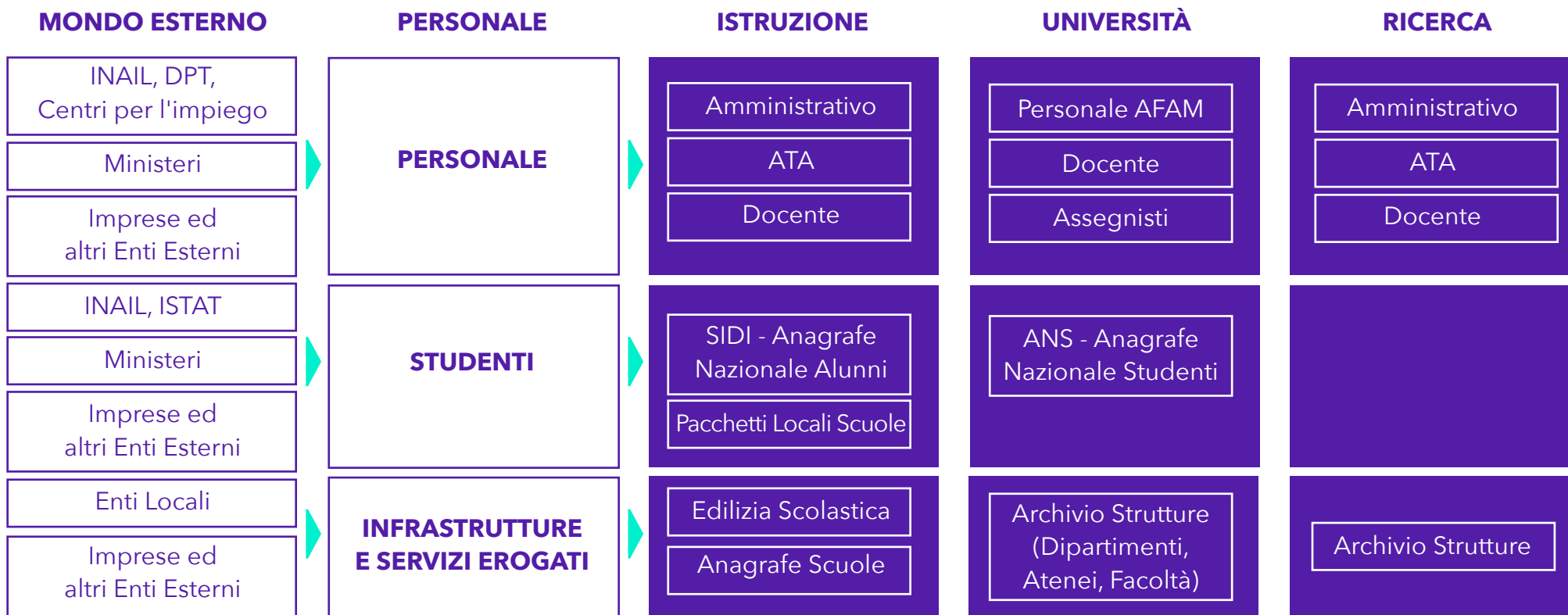
**Il sistema informativo della RICERCA** è classificabile in due macro-aree. I dati che le descrivono sono rispettivamente:

- **Progetti e finanziamenti:** progetti finanziati dal Fondo di Incentivazione della Ricerca di Base (FIRB), Programmi di Ricerca Scientifica di Rilevante Interesse Nazionale (PRIN), progetti e relativi importi e beneficiari (PONREC), procedure amministrative per il finanziamento di progetti di ricerca industriale e innovazione (SIRIO), risultati dei progetti di ricerca e innovazione conclusi (SCRISAO).
- **Strutture e soggetti:** dati relativi agli Enti della Ricerca scientifica e delle Università ammessi e non ammessi al beneficio (5 per mille), ai laboratori di ricerca, agli istituti ed agli enti pubblici e privati che hanno beneficiato e/o partecipato a bandi di ricerca o a procedure selettive di finanziamenti pubblici per la realizzazione di attività di ricerca e/o di formazione (ARIANNA).

Macro Area	I Livello	II Livello	Responsabilità del dato
Ricerca	Progetti e finanziamenti	SIR	DGRIC
		PRIN	DGRIC
		PONREC	DGRIC
		Sirio - richiesta finanziamenti	DGRIC
		5 per mille	DGRIC
		Scrisao (Risultati della ricerca)	DGRIC
	Strutture e soggetti	Elenco laboratori	DGRIC
		Arianna - anagrafe delle ricerche	DGRIC



**MIUR**





## 2 Proposte

Il patrimonio informativo digitale generato e gestito dalle istituzioni e amministrazioni pubbliche nell'esercizio delle rispettive funzioni costituisce una porzione rilevante, in termini di volumi e varietà, della produzione nazionale di *Big Data*. Il valore di queste informazioni cresce se esse sono aggregate e integrate, anche con altri dati derivanti dal crescente utilizzo del web, dei *social network*, per promuovere nuovi modelli interpretativi e nuova possibile conoscenza al fine di contribuire allo sviluppo di innovativi servizi e funzionalità, verso la cosiddetta *data-driven society*.

Già oggi le politiche nazionali richiamano una forte attenzione verso un processo di valorizzazione del patrimonio informativo istituzionale incentivando l'accesso, il riuso e la condivisione dei dati (vedi disciplina del riuso, Decreto legge 18 maggio 2015, n. 102 - Attuazione della direttiva 2013/37/UE che modifica la direttiva 2003/98/CE, relativa al riutilizzo dell'informazione del settore pubblico).

### **LA GOVERNANCE DEI DATI**

La capacità di governare, sia dal punto di vista gestionale che analitico, queste immense moli di informazioni, necessita che siano definiti alcuni requisiti tecnici per favorire la fruibilità del dato e garantire la legittimità del riuso. Un forte incentivo in tal senso è offerto dalle linee guida nazionali per la valorizzazione del patrimonio informativo pubblico dove, a fronte di una

diffusa gestione dei *repository* istituzionali come "isole di informazioni, con scarsa visione sistemica", sono indicati i presupposti per attivare un processo di produzione e rilascio dei dati standardizzato e interoperabile su scala nazionale.

### **DATA PROTECTION**

Viste le forti implicazioni legali connesse al trattamento dei dati, al fine di superare alcuni ostacoli oggettivi che limitano fortemente i processi di riuso, anche meramente istituzionale, è necessario definire *policy* nazionali sui temi della *data protection*. In tal senso lo sviluppo di soluzioni modulari, che prevedano pubblicazioni e accessi differenziati in base ai ruoli e alle specificità degli obiettivi, rende possibile incentivare i processi di *data analytics*, limitandone i rischi connessi. In particolare, le informazioni su dati sensibili possono essere anonimizzate e quindi deidentificate e desensibilizzate.

### **OPEN GOVERNMENT**

Questo processo di valorizzazione del patrimonio informativo richiede un uso e riuso dell'informazione interna non solo in quanto funzionale al reale soddisfacimento delle istanze sociali di trasparenza e responsabilizzazione della PA in un'ottica di *open government*, in sintonia con altri movimenti e comunità di sviluppo "open", ma anche in quanto strumentale a stimolare *policy* tese al costante miglioramento della qualità dei servizi erogati e dell'efficacia ed efficienza della *performance* istituzionale.

## CAMBIARE MARCIA NELLA PRODUZIONE E ORGANIZZAZIONE DELLE INFORMAZIONI

Similmente a quanto caratterizza molti settori pubblici, anche nell'ambito del MIUR, qui di specifico interesse, l'autonomia che contraddistingue i diversi comparti e le rispettive articolazioni ha contribuito allo sviluppo di distinte modalità di gestione delle informazioni, generando un'eterogeneità di *repository* isolati e difficilmente comunicanti tra loro. Un'informazione isolata presenta un'utilità circoscritta, diversamente la combinazione e aggregazione di dati afferenti ai diversi comparti ministeriali dà origine ad un bacino di informazioni tra loro comparabili, integrabili che ne agevola il controllo e il riuso.

L'implementazione di un ambiente ministeriale centralizzato di *data* e *content management* trova oggi nuovo sostegno nel fenomeno dei *Big Data*. I centri di produzione e di raccolta per la preservazione, la cura e il processing dei dati, anche alla luce delle *best practices* promosse in altri contesti, possono essere oggi sollecitati a considerare il valore offerto dall'interoperabilità dei data set e dalle opportunità generate da processi di *business intelligence* su un unico ambiente che nasca dalla convergenza e integrazione dei molteplici data set istituzionali.

Questi obiettivi inducono un cambiamento nei tradizionali paradigmi di produzione e organizzazione dell'informazione, in termini di sviluppo di nuove tecnologie di conservazione, raccordo, condivisione e trattamento, nonché di gestione della stesse rispetto

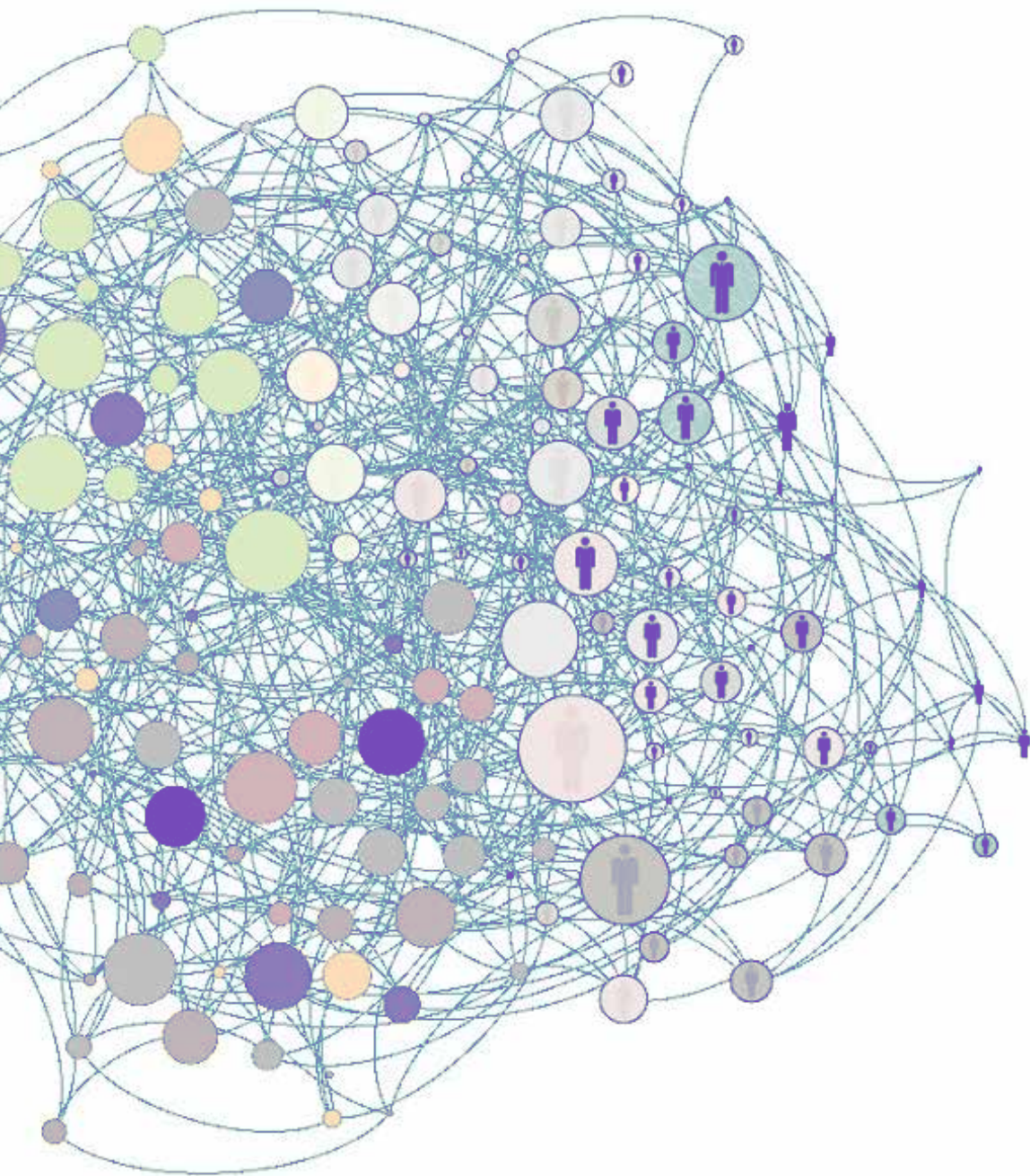
alla celerità con cui i dati necessitano di essere registrati e analizzati.

I nuovi modelli di management offerti dai *Big Data* possono trasformare le procedure ministeriali e aprire nuove frontiere di analisi rispetto a quelle convenzionalmente impiegate in termini di capacità di trattamento della varietà, volume e velocità. I tradizionali modelli di *data mining* sono funzionali a estrarre conoscenza mediante *database* o *data warehouse* e *data mart* secondo un approccio statico e retrospettivo. Diversamente, i processi di *data analytics* di ultima generazione permettono di incrementare le capacità elaborative di straordinari volumi di dati, strutturati (*database*) e non strutturati (file di testo a carattere narrativo oppure file multimediali), costantemente aggiornati, e pertanto rappresentano tecnicamente una grande opportunità per aggregare, analizzare e sfruttare le diverse combinazioni informative ministeriali [13].

L'incremento di tali processi permette di reinterpretare i *dataset* Miur esplicitando il valore insito nella natura eterogenea dei dati collezionati e generare nuove informazioni e nuova conoscenza da investire all'interno della pratica istituzionale e nei rispettivi domini di competenza.

Un ripensamento della gestione delle informazioni in linea con questi innovativi *trend* metodologici, non può prescindere dal rivalutare una parte sottostimata del patrimonio Miur, rappresentata dalle collezioni documentali digitali. A fronte di una gestione dei processi di *business intelligence* (*data integration*, *data*





*warehousing*) che, pur con i limiti imposti da una gestione settoriale delle informazioni, sono maturati in maniera consistente nel corso dell'ultimo decennio, i documenti digitali collezionati dall'apparato ministeriale sono basi informative non ancora sufficientemente esplorate. Il trattamento automatico dei *Big Data* tramite le tecnologie semantiche consente oggi di estrarre e concettualizzare molte delle informazioni contenute nei file di testo dei documenti elettronici collezionati nei *repository* ministeriali (ad es. prove di valutazione, esercitazioni, documentazione progettuale, provvedimenti di rendicontazione depositati sui circuiti ministeriali). Attraverso i traguardi raggiunti da applicazioni *Natural Language Engineering* anche le informazioni nascoste all'interno dei testi (*Big Data* testuali) possono essere esplicitate e assumere utilità diverse rispetto ai fini istituzionali per i quali i documenti sono stati prodotti o raccolti e contribuire ad arricchire il patrimonio informativo ministeriale.

**Allo scopo di ampliare l'impatto di queste azioni, nonché di offrire strumenti adeguati a disposizione del sistema educativo si propone che il MIUR, nell'ambito della sua azione di valorizzazione dei propri dati e quelli della ricerca evidenziata soprattutto al paragrafo 2.2 di questo capitolo, gestisca e offra al sistema educativo delle palestre di dati, ossia ambienti espressamente destinati ad accompagnare i poli di formazione e di crescita dei futuri *data scientist*.**

# 2.1 Policy sulla strutturazione e fruibilità dei dati MIUR

Come evidenziato finora, il Ministero dispone di un grande patrimonio di dati, frutto delle procedure amministrative e gestionali, di apposite rilevazioni statistiche, delle attività di valutazione e autovalutazione della didattica e della ricerca.

Le informazioni raccolte sono già oggi utilizzate per la formulazione delle politiche e per l'analisi dei sistemi scolastico, universitario e della ricerca. Sono in parte a disposizione delle singole istituzioni per le loro scelte gestionali, in parte rese accessibili in modo aperto al più ampio pubblico attraverso dei portali dedicati e utilizzate dal Sistema Statistico Nazionale. Ciononostante il patrimonio informativo del Ministero non è ancora pienamente valorizzato né reso fruibile nelle modalità ottimali.

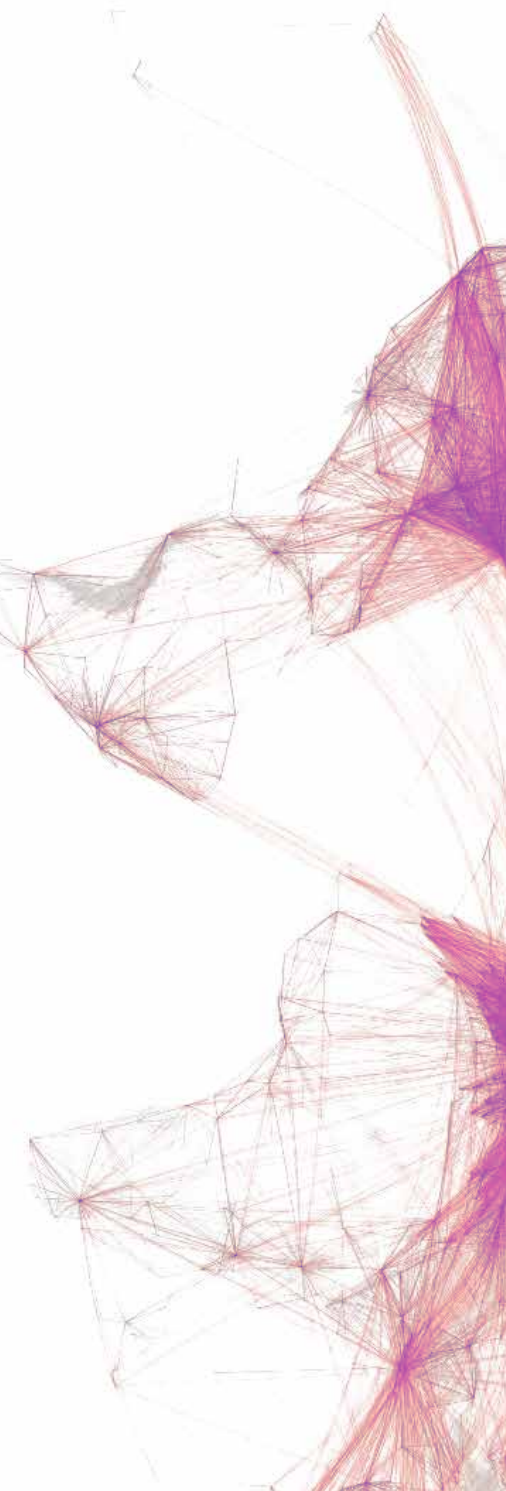
Il MIUR, data la ricchezza dei dati e le esperienze già maturate nel loro utilizzo, è nelle condizioni di rea-

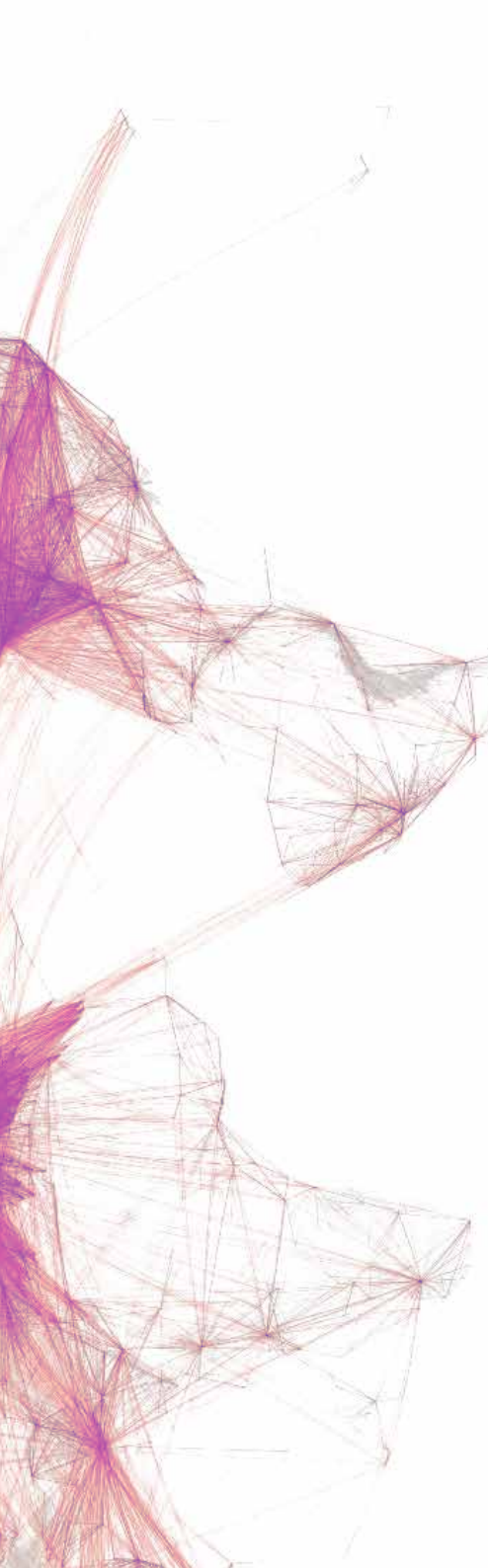
lizzare una **strategia di sistematica valorizzazione dell'informazione e ambire a definire** in questo ambito le migliori pratiche delle amministrazioni centrali. In tempi ragionevolmente brevi può aggiungersi agli esempi di buon governo nella gestione e nell'utilizzo dei dati ai fini della definizione delle politiche e del corretto rapporto con le strutture amministrative e il cittadino.

## **I contenuti, l'organizzazione dei dati e l'integrazione con le informazioni esterne**

Il patrimonio di informazioni del MIUR può essere classificato sulla base dei soggetti a cui le stesse si riferiscono. Semplificando, si possono individuare tre principali categorie:

- 1. Le singole istituzioni** (scuole, università, enti di ricerca e altre istituzioni di alta formazione): informazioni sulle strutture, sull'offerta formativa, sulle risorse economiche e umane, sulla produzione scientifica e quelle derivanti dalle attività di autovalutazione e valutazione.
- 2. Gli studenti:** informazioni relative al percorso formativo e ai principali eventi nella carriera scolastica e universitaria.
- 3. Il personale** docente e di ricerca e il personale amministrativo: informazioni sul percorso professionale e sul trattamento economico, ma anche, in maniera eterogenea per le diverse carriere, sulla loro produzione scientifica, sugli incarichi ricoperti e la formazione.





Questo insieme vasto di informazioni è spesso frutto di distinte procedure amministrative a cui corrispondono differenti responsabilità gestionali e differenti sistemi informativi sottostanti. Emergono così “linee di frattura” nella definizione delle variabili, nella loro codifica, nella gestione informatica dei dati. Queste possono riguardare i soggetti cui le informazioni si riferiscono, la tipologia di amministrazioni coinvolte (scuola, università e ricerca), la stessa organizzazione dei sistemi informativi del MIUR, frutto a sua volta della stratificazione storica di un ministero nato dalla fusione del Ministero dell’Istruzione con quello dell’Università e della Ricerca.

#### **DOCUMENTARE E CODIFICARE**

Vi è quindi la necessità prioritaria di intraprendere un’azione di organizzazione del patrimonio informativo che garantisca l’interoperabilità dei diversi *data base*, indipendentemente dai processi amministrativi da cui originano, delle amministrazioni a cui si riferiscono, dei sistemi gestionali adottati. L’obiettivo di una riorganizzazione del sistema informativo, con la definizione di ontologie comuni e di un sistema di metadati coerente che dia la possibilità di connettere tutte le informazioni disponibili, dovrebbe essere guidato dall’obiettivo di assicurare lo sfruttamento informativo dei dati disponibili. Ciò significa **garantire che ogni dato raccolto nell’ambito di un’apposita procedura amministrativa sia accuratamente documentato e codificato in maniera omogenea nell’intero sistema informativo integrato.**

#### **I DATI DEL MIUR SONO GRANDI MA NON BASTANO**

Le informazioni direttamente gestite dal Ministero non esauriscono le sue esigenze conoscitive né quelle delle singole strutture su cui il MIUR esercita il suo governo. Dovrebbero quindi essere integrate, laddove necessario, con dati provenienti da altre amministrazioni e con i dati raccolti dall’Istituto Nazionale di Statistica. A titolo di esempio, ulteriormente analizzato in seguito, si possono citare i dati sulla condizione occupazionale di quanti escono dal percorso formativo. Tali dati sono disponibili presso altre amministrazioni e possono essere collegati a specifici percorsi formativi per valutarne gli esiti lavorativi. Tuttavia, alla luce del loro assoluto interesse per l’orientamento delle scelte formative e come strumento di analisi dell’efficacia dei percorsi scolastici e universitari, dovrebbero essere **integrati in un singolo data set immediatamente disponibile per ogni indagine sia interna del MIUR sia di altri enti**, ferma restando l’esigenza di assicurare i necessari profili di *privacy*.

Le informazioni generate in altri ambiti arricchiscono e accrescono il valore e l’utilità dei dati già a disposizione, rispondendo a un fabbisogno conoscitivo che non dovrebbe trovare barriere nella responsabilità amministrativa dei diversi Ministeri e organi di governo. Per rimanere sul caso delle informazioni sulla condizione occupazionale, **va rafforzato un approccio unitario superando una visione secondo la quale, ad esempio, il MIUR gestisce e analizza i dati di scuola e Università e il Ministero del Lavoro si occupa dei dati sull’occupazione e delle politiche per l’occupazio-**

**zione, laddove i due ambiti necessitano di un chiaro e sistematico collegamento.** Azioni in questo senso sono già state intraprese, ma qui si sottolinea la necessità di una completa integrazione di queste informazioni in un **singolo set di dati direttamente interrogabile senza intermediazioni.** Altri attori possono poi essere inclusi in questo percorso di integrazione come, ad esempio, l'Istat, che può utilmente partecipare non solo come utilizzatore di dati per la produzione della statistica ufficiale, ma anche come fornitore di dati alle singole amministrazioni per l'analisi dei fenomeni e l'elaborazione di policy arricchendo i loro dati specifici con il patrimonio informativo da esso raccolto presso soggetti istituzionali, con le proprie indagini o estratto dal web. Infatti, **oltre all'informazione organizzata disponibile presso le amministrazioni, oggi è possibile sfruttare in una logica Big Data anche le informazioni liberamente ricavabili appunto dal web e dai social network.** Inoltre, le consultazioni on line sulle scelte di *policy* possono utilmente integrare le informazioni ricavabili dai dati strutturati. Questo insieme di informazioni può velocemente essere raccolto per valutare processi in corso, che nel caso del Ministero possono coinvolgere decine di migliaia di persone. A titolo di esempio, il MIUR nel mondo della scuola intraprende azioni di formazione di massa sulle quali informazioni ricavate dai social network e dal web possono fornire ai decisori un immediato riscontro sulla qualità percepita dai diretti interessati ed eventualmente adottare i necessari correttivi. A loro volta queste informazioni potrebbero essere integrate con i dati strutturati, traendo il massimo beneficio dall'insieme di informazioni raccolte.

## I MOLTI DESTINATARI DEI BIG DATA DEL MIUR

Le informazioni direttamente gestite dal MIUR, eventualmente integrate da informazioni esterne delle diverse fonti, dovrebbero essere rese accessibili e organizzate sulla base delle diverse esigenze conoscitive che queste possono soddisfare. Si possono individuare diversi destinatari:

- 1. il decisore politico**, nello specifico il Ministero e gli organismi preposti alla valutazione, direttamente coinvolti nella gestione, nella formulazione delle politiche e nella valutazione degli esiti;
- 2. le singole amministrazioni** governate dal Ministero (le scuole, le università, gli enti di ricerca), che dalla disponibilità di dati sulla propria e sulle altre strutture possono trarre le necessarie indicazioni per le scelte gestionali e di orientamento dell'attività didattica e di ricerca;
- 3. il sistema statistico nazionale** che utilizza le informazioni statistiche prodotte dalle amministrazioni sui diversi ambiti della società e dell'economia;
- 4. il mondo della ricerca**, che in maniera indipendente è interessato all'analisi dei percorsi di studio e professionali degli studenti, e all'analisi della *governance* del sistema scolastico universitario e della ricerca;
- 5. i singoli cittadini e le famiglie** che hanno interesse ad acquisire informazioni sul sistema scolastico e universitario;

**6. il mondo degli sviluppatori** di applicazioni, piattaforme e servizi a contenuto informativo, che possono costruire a partire dalle informazioni pubbliche del Ministero nuove soluzioni a carattere sociale e commerciale.

I diversi "attori" hanno esigenze conoscitive non omogenee. Allo stesso tempo l'accessibilità dei dati deve essere differenziata sulla base del rispetto delle normative a tutela della privacy dei diretti interessati: si pensi alle informazioni sulle carriere ed esiti negli studi dei singoli studenti. **Ciò richiede un'organizzazione dell'informazione e un sistema di policy di accesso e riuso dei dati che possa rispondere alle diverse esigenze conoscitive nel pieno rispetto dei ruoli e delle responsabilità.**

*Ad esempio, i cittadini hanno particolare interesse ad avere una rappresentazione articolata del piano dell'offerta formativa nel loro contesto territoriale, anche in termini della efficacia degli istituti, laddove il decisore politico e le amministrazioni hanno la necessità di incrociare diverse tipologie di dati per avere una valutazione complessiva dell'impatto delle loro scelte politiche e gestionali e poter velocemente effettuare analisi di tipo predittivo per poterne valutare gli effetti anche a lungo termine. D'altro canto gli organi di governo del Ministero devono poter accedere direttamente - e in tempi rapidi - a tutta l'informazione disponibile, sia di fonte interna sia esterna, per poter rispondere rapidamente alle esigenze conoscitive che di volta in volta si pongono alla loro attenzione. Il lavoro intrapreso in questo senso attraverso i "cruscotti" EduReport e RAV è un passo utile in questa direzione e merita di essere completato e rafforzato.*

**Tutto questo richiede un investimento organizzativo, in modo che sull'insieme di soggetti individuati (strutture, studenti, personale) siano sempre accessibili i dati rilevanti, tra loro interconnessi. Vi è la necessità di costruire a priori un set informativo minimo, in cui tutti i dati di maggior rilievo siano rapidamente recuperabili e allo stesso tempo dotarsi della possibilità di svolgere analisi più in profondità con il diretto ricorso alle informazioni elementari.**

## UNA VISIONE COORDINATA A PARTIRE DA DIVERSE FONTI DI DATI

Le singole strutture governate dal Ministero (scuole, università, enti), spesso **fonte primaria del dato**, devono essere messe nella condizione di sfruttare le informazioni a disposizione del MIUR per le loro scelte gestionali e di indirizzo strategico, potendosi confrontare su base quanto più possibile omogenea con le altre. A titolo di esempio, l'ANVUR e le stesse università hanno la necessità di confrontare i percorsi di carriera degli studenti nei diversi atenei, in modo da poter misurare l'efficacia delle scelte didattiche e organizzative.

Il Sistema Statistico Nazionale (SiStaN) utilizza sempre di più informazioni fornite dalle singole amministrazioni, generate da specifici processi amministrativi. Ciò garantisce una vasta copertura dell'informazione e permettere di ridurre il ricorso alle indagini campionarie allo stretto indispensabile. Il MIUR già fornisce una mole rilevante di dati, svolgendo una funzione di primaria importanza per la costruzione di un quadro statistico affidabile sulla formazione dei cittadini italiani.

Allo stesso tempo il MIUR potrebbe certamente beneficiare da appositi accordi per dare maggior completezza all'insieme di dati di cui dispone, ricevendo dall'ISTAT informazioni in suo possesso necessarie per lo svolgimento dei compiti istituzionali del Ministero. Si attuerebbe così un **processo di interazione e di scambio, orientato allo sfruttamento a fini conoscitivi e decisionali dei dati a disposizione** che potrebbe fornire sia al Ministero che all'ISTAT indubbi van-

taggi in termini di assicurazione della qualità del dato e sua valorizzazione. Nella stessa direzione, al fine di garantire la completezza delle informazioni, sarebbe auspicabile l'apertura di un tavolo di lavoro in ambito SiStaN che preveda anche la partecipazione di Invalsi e Anvur.

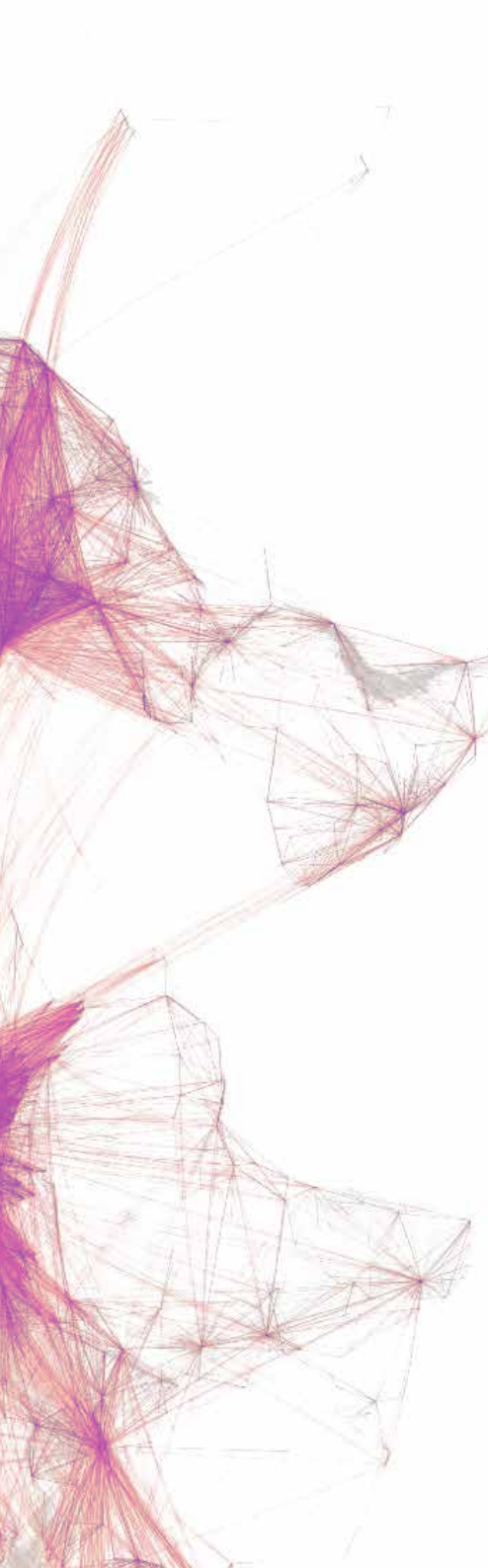
## Il MIUR può fornire al mondo della ricerca un ampio insieme di informazioni sul sistema dell'istruzione e della ricerca.

Ciò può avvenire in forma aggregata, come già in parte avviene, ma anche prevedendo la possibilità di accedere a dati individuali, garantendo i necessari standard di tutela della privacy. **L'apertura al mondo della ricerca può alimentare un dibattito scientifico competente e articolato sui temi di interesse diretto del Ministero.** Ciò stimolerebbe analisi e riflessioni sull'impatto delle scelte di policy che a loro volta potrebbero fornire strumenti per lo sviluppo di processi decisionali ancor più informati e scientificamente supportati, anche a beneficio dell'opinione pubblica. Occorre definire un insieme di **informazioni standard** da fornire in maniera regolare all'esterno, integrando e riorganizzando il materiale già a disposizione. Allo stesso tempo il MIUR potrebbe utilmente consentire l'**accesso controllato ai dati micro**, che, con le dovute garanzie di tutela della privacy, potrebbero permettere studi scientifici indipendenti con lo *standard* qualitativo della migliore letteratura internazionale.

Infine il Ministero, come in parte già realizzato con la piattaforma **Scuola in Chiaro** e con le informazioni dif-







fuse tramite il sito **University**, dovrà rendere fruibili le informazioni alle famiglie e agli studenti, attraverso piattaforme o servizi (API) che permettano il pieno accesso ai dati pubblici, secondo modalità tecnicamente allo stato dell'arte e che rispondano, anche graficamente e in relazione ai profili della *user experience*, sia alle esigenze del pubblico non specialista, sia degli utenti che intendano riutilizzare i dati per scopi di ricerca o commerciali.

È attualmente in fase di elaborazione il progetto del **Portale Unico dei Dati Scuola**, pensato per favorire un'agevole lettura dei dati. Questo garantirà un elevato livello di trasparenza delle strutture su cui il MIUR esercita la sua azione di governo e accrescerà il grado di consapevolezza dei cittadini nelle scelte di istruzione. Le informazioni disponibili sulla sua strutturazione fanno ritenere questo un obiettivo importante e utile di cui si auspica una sua realizzazione che, come già menzionato più sopra in altro contesto, possa beneficiare anche del contributo di altri enti quali INVALSI, ANVUR, INDIRE e ISTAT.

#### **L'AZIONE INTERNA PER LO SFRUTTAMENTO DEI DATI**

Dal punto di vista operativo, sono qui proposte alcune linee strategiche generali e specifiche azioni che possono essere intraprese a breve e medio termine per garantire l'interconnessione delle basi dati MIUR e il miglioramento della loro fruibilità interna rispetto ad analisi avanzate. Dalla ricognizione effettuata emerge che il Ministero ha già implementato alcune di queste linee. Tuttavia, per fornire un quadro coerente e il più

possibile completo verranno esplicitate anche linee già presenti nell'operatività attuale del MIUR.

Il punto di partenza dell'attività è il **censimento delle basi dati del MIUR**, che deve includere le attuali modalità informatiche di implementazione degli archivi e del loro accesso, nonché il loro stato di aggiornamento e di documentazione con il fine di assicurare:

- che ogni base di dati sia memorizzata mediante un supporto tecnico adeguato (sistema di gestione di basi di dati relazionali) e sia fornita di un progetto ben documentato;
- l'identificazione delle entità primarie dei vari sistemi informativi (studenti, scuole, ecc.) e la loro codificazione in maniera coerente, al fine di rendere possibile un collegamento tra i diversi archivi;
- l'implementazione di processi di aggiornamento continuo laddove i dati risultino non idonei in termini di tempestività e coerenza degli aggiornamenti.

Si suggerisce una strategia di approccio *bottom-up* incrementale che possa garantire risultati a breve termine, rilasciati con continuità e frequenza, convergendo allo stesso tempo su una visione integrata di lungo termine. Partendo da interventi mirati su problemi specifici a più alto impatto strategico e/o economico, l'obiettivo finale deve essere quello di riconciliare le informazioni presenti nelle basi di dati e fornire una piattaforma integrata basata su tecnologie avanzate che contribuiscano in diversi modi a implementare la

politica di gestione dei dati: un'utile base per questo lavoro è rappresentata dalla piattaforma EduReport. Attraverso questo lavoro i benefici attesi sono molteplici:

- l'aumento della qualità complessiva dei dati, garantendo tempestività degli aggiornamenti, eliminazione delle ridondanze e riconciliazione dei dati presenti in diversi archivi;
- una maggiore fruibilità dei dati per l'utenza interna del ministero, con la possibilità di scaricare facilmente selezioni di dati che possono essere utilizzati per analisi specifiche effettuate da singoli utenti;
- la possibilità di realizzare cruscotti interattivi che forniscano viste navigabili e personalizzabili sui dati relativamente ai principali indirizzi strategici del MIUR, da destinare alle diverse tipologie di utenza sopra identificate;
- la possibilità di effettuare analisi di tipo predittivo che possano supportare in maniera sostanziale e tempestiva le politiche governative;
- Il rilascio di parte dei dati e delle analisi in modalità aperta (e linked).

Complessivamente l'intervento deve essere **orientato a incrementare la cultura del dato, cioè una capacità di fruizione ed elaborazione non legata esclusivamente a un'utenza specialistica ma che investa l'attività del ministero a tutti i livelli**, migliorando la

produttività individuale e l'efficacia complessiva nella programmazione strategica. In dettaglio, si propone di considerare le seguenti linee di attività, suddivise in **attività a breve termine** (alcune delle quali peraltro già avviate), orientate a ottenere risultati concreti su obiettivi specifici, e **attività a medio termine**, orientate a far convergere i risultati delle singole attività verso una visione unificata e stabile.

#### **Linea di attività a breve termine:**

- Interventi di adeguamento di qualità sulle basi di dati esistenti e definizione delle politiche di manutenzione e aggiornamento;
- Raccolta dei requisiti con identificazione dei settori strategici in cui il fabbisogno informativo non è soddisfatto dall'offerta attuale;
- Ricognizione dei dati di altri enti nazionali (Alma-Laurea, Ministero del Lavoro, Istat, INPS) e statistiche internazionali e progetto delle modalità di raccordo e integrazione con le basi dati MIUR;
- Definizione dell'architettura tecnologica e progettazione delle modalità tecniche di integrazione a livello dati e applicazioni;
- Identificazione e realizzazione dei primi interventi di integrazione sui contesti critici identificati nella raccolta requisiti.





## Linea di attività a medio termine:

- Creazione di un livello semantico di ontologie che permetta di integrare tutte le informazioni, sia presenti nelle basi dati ministeriali che ottenute dalle fonti esterne. Questa attività si sostanzia nella definizione formale di un vocabolario di termini e delle relazioni tra essi, con l'identificazione dei sinonimi e dei concetti assimilabili tra loro e la creazione della corrispondenza con gli elementi presenti nelle basi dati. In questo modo si crea un livello standard di definizione della semantica dei dati che consente di facilitare le interrogazioni e le analisi, garantendo una più alta qualità del dato e una manutenzione più efficace.
- Realizzazione di un deposito unico in cui far confluire tutti gli archivi integrati e le fonti dati ausiliarie, anche non strutturate (ad esempio quelle provenienti da web). Questo componente, implementato mediante tecnologie allo stato dell'arte, permetterà di velocizzare l'accesso ai dati e facilitare analisi complesse che sappiano collegare sia i dati nei formati più strutturati e standardizzati di origine interna del MIUR, sia quelli esterni. All'interno del gruppo di lavoro è stato anche ritenuto valido un approccio alternativo, che prevede accordi e motori di ricerca avanzati, tale integrazione verrebbe realizzata direttamente sul web.

## POSSIBILI ANALISI TRASVERSALI: ALCUNI ESEMPI

Le attività descritte renderanno possibili analisi di tipo trasversale tra le varie basi di dati, non possibili nel contesto attuale, di cui si forniscono alcuni esempi nel seguito:

- **Studenti:** seguire il percorso della carriera scolastica di un singolo studente tra scuola primaria, media e università e calcolare indicatori avanzati di performance dei diversi percorsi formativi. Questo caso d'uso verrà dettagliato nella sezione seguente.
- **Scuola:** incrocio tra i dati di performance ottenuti dal sistema di autovalutazione e il quadro dei finanziamenti ordinari e a carattere progettuale a cui le istituzioni scolastiche accedono, al fine di effettuare una valutazione dell'efficacia degli investimenti, anche rispetto al territorio, e identificare meglio i fabbisogni e le aree di criticità.
- **Povertà educative, periferie e abbandono scolastico:** applicazione di tecniche di machine learning per identificare e classificare i fattori più rilevanti rispetto ai fenomeni di abbandono scolastico, efficacia dei percorsi educativi anche in aree non geograficamente definite, quali le "periferie". Si possono a questo scopo integrare le informazioni relative alle carriere degli studenti, incluse attività didattiche ed extracurricolari, con altri dati che possono definire il contesto socio-demografico e applicando algoritmi automatici di classificazione si può dare una caratterizzazione più precisa degli aspetti che più influiscono

sull'abbandono e la povertà educativa, in modo da poter definire delle politiche specifiche a riguardo.

- **Studenti e mercato del lavoro:** Valutare l'offerta formativa su uno specifico territorio rispetto all'offerta professionale nei vari settori. Si possono utilizzare fonti ausiliarie come dati del ministero del lavoro e dati raccolti dai portali web degli annunci di lavoro, per realizzare analisi che caratterizzando l'offerta di lavoro sul territorio permettano di identificare i contesti territoriali dove questa non viene adeguatamente supportata dall'offerta formativa.

I risultati delle analisi potranno essere resi disponibili anche come dati aperti, fruibili da ricercatori e cittadini.

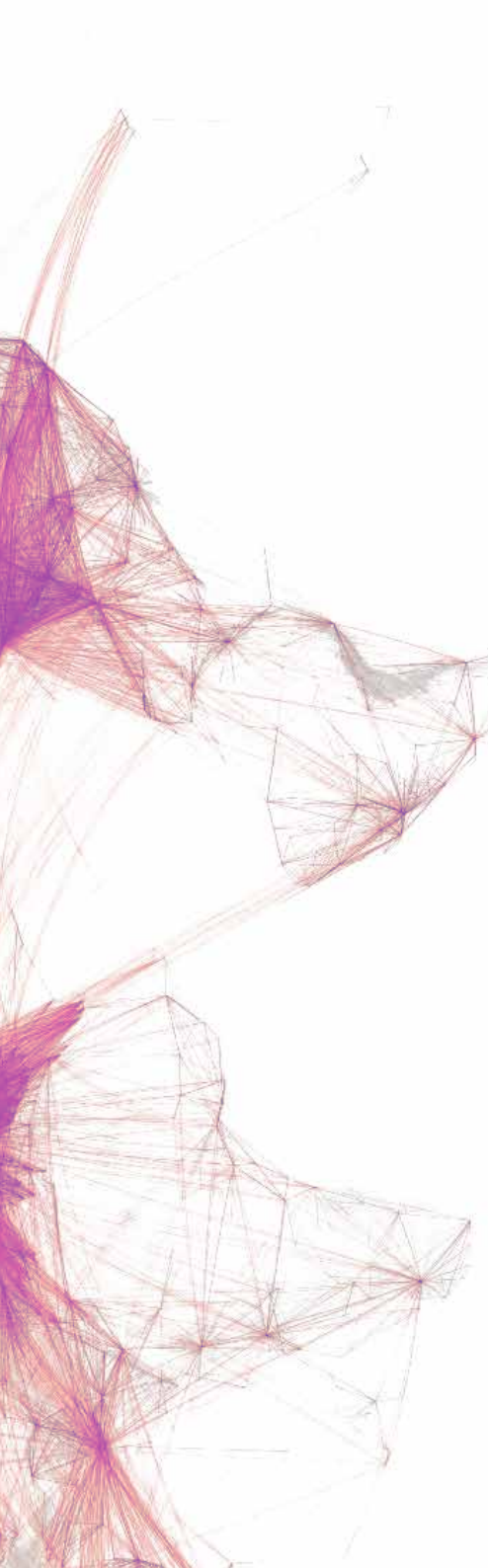
### **POTENZIARE L'INFRASTRUTTURA DEL MIUR**

Una strategia di sfruttamento dei dati con tali ambizioni richiede la valorizzazione delle strutture preposte alla gestione e soprattutto all'analisi dei dati. **Devono essere potenziate le capacità analitiche, in modo da assicurarne l'efficace supporto al processo decisionale.** Alle stesse strutture deve essere garantita la possibilità di definire, in collaborazione con quelle con responsabilità di gestione, gli *standard* di interoperabilità delle banche dati e garantire la loro accessibilità diretta da parte del Ministero e degli altri organi di governo del sistema scolastico, universitario e della ricerca. È inoltre necessario **un progetto che delinei le singole azioni di sfruttamento dei dati a disposizione secondo le diverse esigenze conoscitive** sopra delineate.

Un orientamento allo sfruttamento dei dati a fini conoscitivi, sostenuto dal necessario impegno organizzativo, permetterebbe al MIUR di accrescere la qualità dei dati e il livello complessivo di consapevolezza delle diverse strutture amministrative sulle potenzialità offerte dal loro sistematico utilizzo nel processo decisionale. Il coinvolgimento delle strutture preposte alla gestione e analisi dei dati nel processo decisionale favorirebbe la maturazione di conoscenze specifiche al servizio delle strutture di governo.

La struttura preposta dovrebbe poter contare su collaborazioni scientifiche esterne e basarsi sulle esigenze conoscitive dei diversi soggetti e strutture, al fine di predisporre un'efficace architettura del sistema e la definizione di set di informazioni *standard* da rendere automaticamente disponibili ai diversi soggetti. Ugualmente, **l'interazione con le strutture tecniche e con la comunità scientifica** permetterebbe una miglior definizione delle informazioni da usare e delle tecniche di analisi più appropriate da utilizzare. In termini di risorse umane, sarebbe necessaria una struttura informatica di servizio che predisponga gli ambienti per l'accesso ai dati, con caratteristiche differenziate a seconda delle tipologie di utenza, siano esse le strutture di governo del ministero, le strutture amministrative o le famiglie, nell'ambito di un progetto unitario. Occorrerebbe inoltre individuare competenze per lo sfruttamento delle informazioni non strutturate disponibili sul web o sui social network, che sappiano indicare le soluzioni più appropriate per l'uso congiunto di informazioni soft e non strutturate e di quelle già a disposizione del ministero.





Infine è necessario che le strutture con compiti di gestione e analisi dei dati del Ministero e tutte le strutture di governo centrali del sistema scolastico, universitario e della ricerca (ANVUR, INVALSI, INDIRE) sviluppino e mettano a disposizione competenze analitiche forti, che oltre alla gestione del dato, pongano enfasi sulla sua interpretazione ai fini della definizione delle politiche.

### **UN ESEMPIO CONCRETO DI SFRUTTAMENTO DEI DATI: LE CARRIERE DEGLI STUDENTI**

Il MIUR gestisce l'Anagrafe studenti delle scuole primarie, secondarie di primo e secondo grado e l'Anagrafe nazionale degli studenti universitari: è quindi nella condizione di seguire il percorso completo degli studi per quasi l'intero universo degli studenti.

I dati raccolti registrano tutti gli eventi rilevanti. Il MIUR è così in grado di monitorare anno dopo anno la regolarità degli studi, gli abbandoni, i cambiamenti del corso di studi nonché i risultati: voti scolastici, numero di crediti e voti conseguiti nei corsi di laurea universitari, votazione finale nelle lauree triennali e magistrali. Queste informazioni possono essere integrate con dati sulle istituzioni scolastiche (ad esempio risultati dei test Invalsi) e quelle universitarie (ricchezza dell'offerta formativa, risorse, valutazioni della qualità di ricerca dei singoli dipartimenti ecc.).

Già oggi il MIUR è in grado di collegare questi archivi, anche se gestiti secondo modalità e architetture molto diverse, e già oggi queste informazioni sono utilizza-

te dagli organi di governo del Ministero (ad esempio nella definizione dei finanziamenti agli atenei), nonché in parte sfruttate a fini di ricerca e analisi dall'ANVUR e da istituzioni di ricerca esterne. Le scuole, le famiglie e i cittadini possono accedere ad alcune informazioni tramite il progetto Scuola in chiaro e attraverso il Sistema Nazionale di Valutazione. ALMALAUREA raccoglie inoltre i dati relativi agli esiti didattici e occupazionali degli studenti universitari. L'ANVUR pubblica nei suoi rapporti biennali dati sulla regolarità degli studi universitari a livello di singolo ateneo. Recentemente le singole università sono state messe nelle condizioni di confrontare la performance dei propri studenti a livello di singolo corso di studio con quella degli altri atenei attraverso un sistema standardizzato di indicatori definito dall'ANVUR, presto sistematicamente calcolabili e resi disponibili tramite una piattaforma sviluppata dall'Ufficio di statistica del MIUR e dal consorzio CINECA.

Questo insieme di dati può essere a sua volta integrato con informazioni esterne al Ministero e gestite da altre amministrazioni, in particolare per quanto riguarda gli esiti occupazionali. Non si tratta di una possibilità astratta: il MIUR ha già sperimentato l'aggancio dei dati individuali sugli studenti medi con i dati del Ministero del lavoro sulle Comunicazioni obbligatorie che, limitatamente al lavoro dipendente, registrano tutte le assunzioni e le cessazioni di un rapporto di lavoro. Sono in corso sperimentazioni per il collegamento con i dati INPS che dispongono delle informazioni sulle retribuzioni. È in corso poi un progetto organico con ISTAT

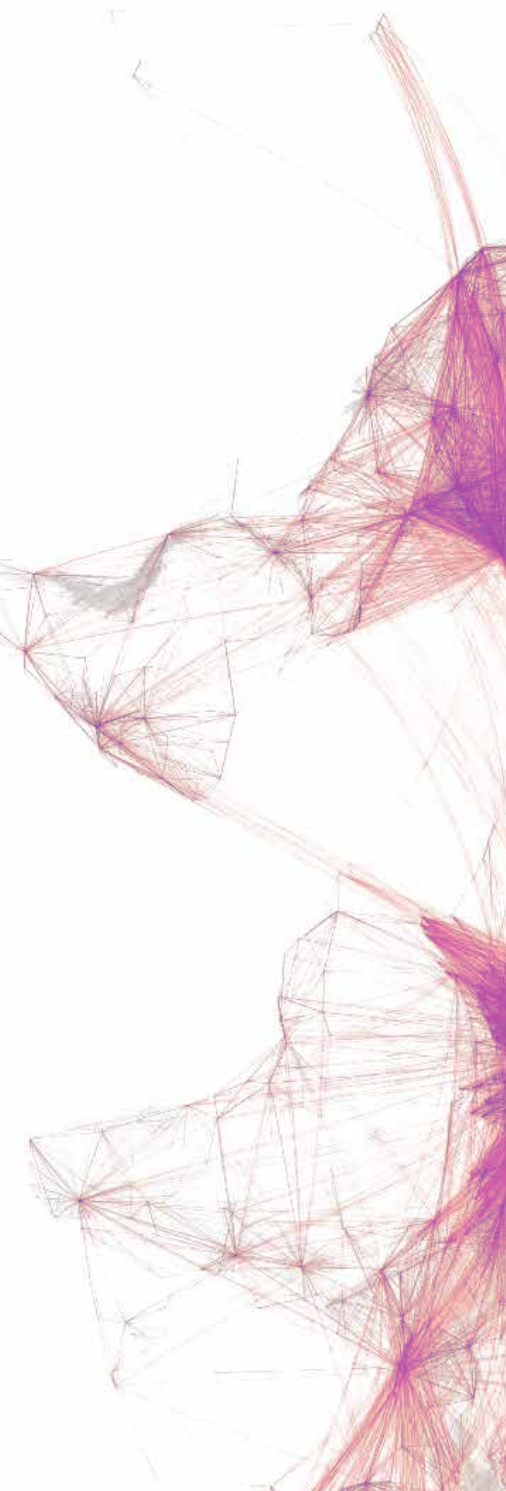
per l'aggancio delle informazioni sui percorsi di studio universitari con i dati amministrativi raccolti dall'ISTAT stesso e che includono i dati Inps e quelli del ministero del lavoro, integrato con i dati delle indagini condotte dal consorzio Almalaurea sui laureati.

Questo esempio mostra le possibilità di sfruttamento e valorizzazione dei dati del MIUR: vi è una chiara utilità dei dati per la definizione delle politiche (si pensi al contrasto della dispersione scolastica); sono dati che possono essere restituiti alle singole strutture per le loro scelte gestionali; possono essere valorizzati integrandone il contenuto informativo con i dati di altre amministrazioni (condizione occupazionale); possono essere utilizzati a fini scientifici, ad esempio per studiare i percorsi formativi e l'inserimento nel mondo del lavoro; possono essere messi a disposizione delle famiglie e dei cittadini in forme già oggi sperimentate.

Date le competenze e le esperienze già maturate nonché la rilevanza e ricchezza di questi dati questo può diventare un progetto bandiera del Ministero su cui sperimentare il percorso di valorizzazione dei dati qui delineato.

Il MIUR può velocemente promuovere una sistematica integrazione dei dati micro sui percorsi scolastici e universitari, arricchiti con le informazioni rilevanti sulle singole strutture. A loro volta le carriere individuali dovrebbero essere integrate con i dati sulla condizione occupazionale attraverso l'evoluzione delle convenzioni già in atto in un progetto unitario che permetta di monitorare l'inserimento lavorativo degli studenti e gli esiti a più lungo termine.

I dati così predisposti dovrebbero essere organizzati e restituiti con modalità specifiche ai singoli potenziali utilizzatori, con piattaforme dedicate agli organi di governo centrale, alle singole strutture governate, alla comunità scientifica e di governo di altre amministrazioni, prevedendo la possibilità di sviluppare progetti scientifici sui dati elementari opportunamente anonimizzati, alle famiglie e ai cittadini secondo modalità già in parte sperimentate.





## 2.2 Policy per la valorizzazione dei dati pubblici e privati

L'interesse verso una massimizzazione del valore dei dati detenuti o prodotti dal settore pubblico deriva dalla strategicità che essi possono rivestire nello stimolo di nuove linee di mercato e di offerta di prodotti e servizi di nuova generazione e con profondo impatto di innovazione sociale.

In tal senso, si possono riconoscere due grandi insiemi di dati che afferiscono a questo ambito: da una parte quelli prodotti e detenuti dal settore pubblico per garantire la corretta e ordinata amministrazione, e d'altra parte, quelli che, finanziati attraverso risorse pubbliche, sono di per sé destinati ad aumentare la base scientifica e di conoscenza del Paese.

Fanno parte del primo insieme i dati pubblici che storicamente sono associati al principio di trasparenza della pubblica amministrazione e che nel tempo si sono evoluti nell'offerta proattiva di banche di dati grezzi [16], generalmente pubblicati seguendo linee guida inter-

nazionali che garantiscono qualità e interoperabilità, e la creazione di infrastrutture digitali per la gestione e l'offerta di dati attraverso reti integrate, soprattutto in relazione a settori informativi riguardanti aree di particolare interesse sociale, quale quella ambientale [17].

Alla seconda categoria si può ricondurre l'accesso ai risultati della ricerca, la loro ampia circolazione e il più efficace utilizzo, preconditione per un adeguato sviluppo di una società innovativa e della conoscenza. Perché la base scientifica pubblica sia pienamente efficace in termini di aumento d'innovazione e d'impatto, serve che questa sia ben collegata al settore delle imprese e che tale conoscenza circoli liberamente.

### **I DATI DEL MIUR, PUBBLICA AMMINISTRAZIONE**

La legge "La Buona Scuola" prevede che tra le banche dati espressamente oggetto di pubblicazione a partire dal 2016, vi siano i dati relativi ai bilanci delle scuole, i dati pubblici afferenti al Sistema nazionale di valutazione, l'Anagrafe dell'edilizia scolastica, i dati in forma aggregata dell'Anagrafe degli studenti, i provvedimenti di incarico di docenza, i piani dell'offerta formativa, i dati dell'Osservatorio tecnologico, i materiali didattici e le opere autoprodotte dagli istituti scolastici e rilasciati in formato aperto.

Perseguendo anche un approccio generale, la norma indica inoltre che il Ministero pubblichi i dati, i documenti e le informazioni utili a valutare l'avanzamento didattico, tecnologico e d'innovazione del sistema scolastico.

Sulla base delle regole e standard del portale “normativa”, una sezione dedicata del portale sugli *opendata* della scuola sarà destinato a pubblicare, operando una razionalizzazione, la normativa, gli atti e le circolari adottati dal Ministero. Sentito il Garante per la Privacy, nel portale sarà infine possibile accedere alla parte pubblica dei curricula degli studenti e del portfolio dei docenti.

I metadati generati dall’apertura delle banche dati di cui sopra confluiranno nel portale dati “Gov.it”, come previsto dalle Linee Guida sulla Valorizzazione del Patrimonio Informativo Pubblico, in modo da renderli estensivamente riutilizzabili, a livello nazionale e internazionale.

**La raccomandazione per questo ambito è che il MIUR non solo si faccia promotore della pubblicazione proattiva dei propri dati, adeguatamente documentati e tecnicamente disponibili adottando le più moderne soluzioni, ma che si coordini con altre basi e flussi di dati pubblici, quali quelli generati dalle azioni locali, soprattutto in riferimento al patrimonio informativo generato dalle città intelligenti, e attivi politiche di interazione, analisi e partenariato con il settore privato per generare prodotti e servizi, ad alta intensità informativa.**

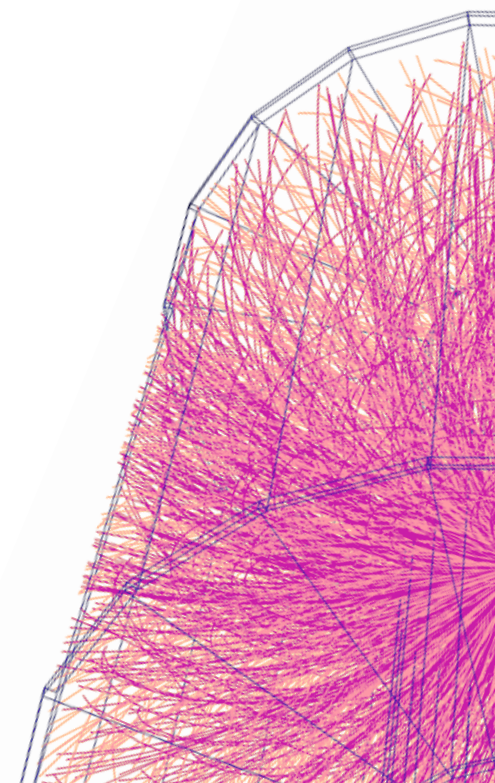
## I DATI DELLA RICERCA: OPEN SCIENCE

Alla seconda categoria di dati si può ricondurre l’accesso ai risultati della ricerca. La loro ampia circolazione e più efficace utilizzo sono considerati precondizioni necessarie per un adeguato sviluppo di una società innovativa e della conoscenza, posto che il loro impatto innovativo è chiaramente collegato alla condizione che essi siano accessibili al settore delle imprese e che la conoscenza circoli liberamente.

Il discorso sull’apertura dei risultati della ricerca è radicato nel tempo: partito da movimenti sulla condivisione aperta delle pubblicazioni scientifiche (Open Access), il dibattito si è evoluto verso la condivisione dei dati primari della ricerca (Open Science) e del sistema dell’innovazione (Open Innovation) [18].

Temi complessi e ibridi come l’Open Science investono ambiti, competenze e problemi assai vasti e articolati che coinvolgono direttamente la governance sia degli Enti di ricerca e delle Università sia del Ministero stesso. Coinvolgono le politiche della ricerca, le scelte strategiche nell’istruzione e formazione, la valutazione, i modelli e le politiche dell’informazione e comunicazione scientifica.

La massima apertura dei risultati della ricerca realizza *prima facie* un obiettivo di efficienza delle risorse pubbliche: l’uso di dati prodotti da una ricerca finanziata precedentemente permette al ricercatore successivo di partire da un punto più avanzato, minimizza poten-





ziali duplicazioni, e quindi ottimizza l'uso delle risorse investite nel sistema della ricerca.

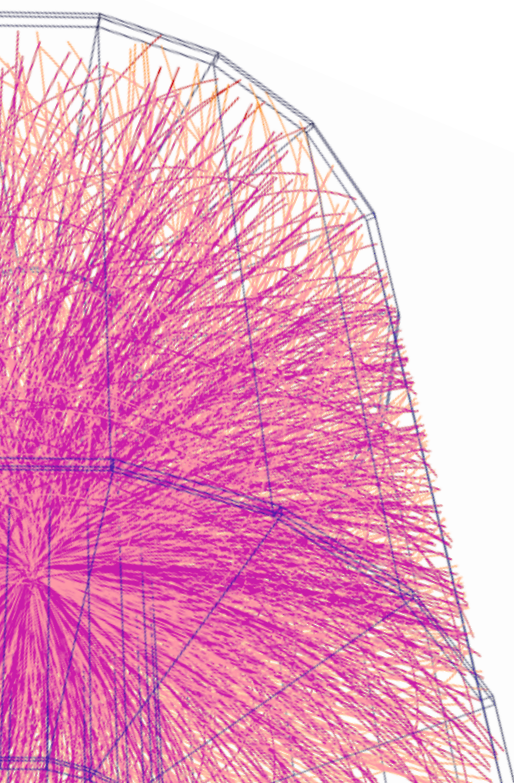
**Il riuso dei prodotti dalla ricerca valorizza quindi il ruolo della comunità scientifica come grande infrastruttura cognitiva fondata sulla collaborazione.**

In secondo luogo, la disseminazione dei risultati della ricerca migliora le relazioni fra scienza (accademia e ricerca) e società, accrescendo da un lato l'impatto e i benefici della R&S sulla società stessa e dall'altro aumentando consapevolezza e fiducia dei cittadini. Viene così massimizzato il ritorno dell'investimento pubblico in ricerca, in piena ottica di *responsabilità sociale* della Ricerca e dell'innovazione, rinforzando la relazione tra scienza e società.

La circolazione dei risultati della ricerca, permettendone il loro successivo riuso, fa in modo che questi abbiano maggiore impatto: le statistiche dimostrano che il mercato Europeo della ricerca, pur producendo un grande volume di "prodotti", maggiore di qualsiasi regione del mondo, del sistema Europeo, non ha un impatto proporzionale, anche di tipo commerciale, generalmente presente altrove.

Occorre d'altra parte sottolineare che l'approccio OpenScience porta con sé delle sfide in parte ancora irrisolte, che occorre affrontare sia all'interno del sistema nazionale, sia collaborando alle molte piattaforme europee e internazionali.

- È necessario implementare un **sistema di incentivi per i ricercatori**, a partire dalle procedure di valutazione MIUR sulla produttività dei ricercatori (VQR), **in modo che sia premiato il riutilizzo e la citazione dei dati da loro messi a disposizione.**
- È auspicabile l'emanazione di **bandi competitivi per il finanziamento di progetti di ricerca che abbiano i Big Data sia come metodo, sia come oggetto di studio.**
- Deve essere previsto su un piano di sensibilizzazione sul valore della condivisione dei dati da parte dei ricercatori con modalità che tutelino la proprietà intellettuale.
- Un'altra sfida attiene alla necessità di **sviluppare infrastrutture elettroniche di ricerca** che siano capaci di ospitare grandi moli di dati per la loro condivisione e conservazione nel tempo. Su questo punto, il MIUR deve promuovere un ruolo attivo dell'Italia per collaborare alla European Cloud Initiative cui tutte le nostre eccellenze del settore hanno immediatamente risposto partecipando da subito alle prime fasi del progetto.



È utile ricordare in proposito che l'attuale Programma Nazionale della Ricerca Italiano intende favorire lo sviluppo e la diffusione di Open Science e *Big Data* nel senso qui indicato. La previsione è di adottare una policy nazionale relativa al deposito, l'accesso aperto, la verificabilità e il riuso dei prodotti e dei dati della ricerca, anche in ottica *Big Data*. Al riguardo, un gruppo di lavoro a questo dedicato definirà e proporrà in tempi brevi le strategie, le linee di indirizzo, piani di attuazione e strumenti condivisi a livello inter-istituzionale, europeo e internazionale per realizzare in Italia l'Open Science, l'Open Access e la Digital Preservation.



# Riferimenti Bibliografici

- [1] Cfr. <http://www-01.ibm.com/software/data/bigdata/>
- [2] MGI. *Big Data: The next frontier for innovation, competition, and productivity*, 2011.
- [3] Oracle. Oracle: *Big Data* for the Enterprise, 2013.
- [4] Tech Partnership - SAS. *Big Data Analytics, Assessment of Demand for Labour and Skills 2013-2020*, Ottobre 2014.  
[https://www.thetechpartnership.com/globalassets/pdfs/research-2014/bigdata\\_report\\_nov14.pdf](https://www.thetechpartnership.com/globalassets/pdfs/research-2014/bigdata_report_nov14.pdf)
- [5] EC. *Towards a thriving data-driven economy*, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, 2014.  
<https://ec.europa.eu/digital-single-market/en/towards-thriving-data-driven-economy>
- [6] Office of Science and Technology Policy -- Executive Office of the President. *Big Data* initiative release, 29.03.2012. [https://www.whitehouse.gov/sites/default/files/microsites/ostp/big\\_data\\_press\\_release\\_final\\_2.pdf](https://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf)
- [7] BYTE European project, (*Big Data* roadmap and cross-disciplinary societal Externalities), *Big Data* Policies, 01.03.2014. <http://byte-project.eu/wp-content/uploads/2016/03/D1.2-Big-Data-Policies-FINAL-post-Y1-Review.pdf>
- [8] See <https://www.gov.uk/government/publications/eight-great-technologies-infographics>
- [9] UK Government, *Seizing the Data Opportunity*, 31.10.2013.  
<https://www.gov.uk/government/publications/uk-data-capability-strategy>
- [10] See <http://www.economie.gouv.fr/vous-orienter/industrie/nouvelle-france-industrielle>
- [11] *Big Data: Feuille De Route*, 04.07.2014.  
<http://www.economie.gouv.fr/big-data-feuille-route-en-action>
- [12] Siemens G., Gasevic D., Haythornthwaite C., Dawson S., Shum S.B., Ferguson R., Duval E., Verbert K., and Baker R. S. J. D. *Open Learning Analytics: an integrated & modularized platform*, 2011. Siemens, G., Long, P. *Penetrating the fog: Analytics in learning and education*. *Educause Review*, 46(5), 2011, 30-32.
- [13] Giacalone M. and Scippacercola S., *Il ruolo dei Big Data nelle strategie di apprendimento*, Proceedings. "Innovazione: sfida comune di scuola, università, ricerca e impresa" 30° EDIZIONE - Udine 19, 20 e 21 Aprile 2016 Didamatica2016
- [14] Gutierrez-Santos S., S. Geraniou, E., Pearce-Lazard, D. & Poulouvassilis. A. *Architectural Design of Teacher Assistance Tools in an Exploratory Learning Environment for Algebraic Generalisation*. *IEEE Transactions of Learning Technologies*, 5 (4), 2012, 366-376.
- [15] Ferguson, R.. *Learning analytics: fattori trainanti, sviluppi e sfide*. *TD Tecnologie Didattiche*, 22(3), 138-147. Ferguson, R., *Learning Analytics: drivers, developments and challenges*. *International Journal of Technology Enhanced Learning*, 4,5/6, 2012, 304-317.
- [16] È da oramai un ventennio che il processo legislativo teso alla valorizzazione commerciale dei dati del settore pubblico ha avuto inizio. Nel 1989 la Commissione Europea ha pubblicato delle "Synergy Guidelines" non vincolanti, ma in ogni caso tese a rafforzare la posizione del settore privato nel mercato europeo dell'informazione e a delineare il ruolo degli organi pubblici in materia di approvvigionamento di dati grezzi. Un decennio più tardi, un passo successivo è risultato in un Libro Verde sull'informazione del settore pubblico, in seguito alla quale fu emessa la proposta di direttiva. Nell'ottobre 2001 fu emessa una comunicazione per la creazione di un quadro europeo per lo sfruttamento dell'informazione del settore pubblico, come parte integrante del piano d'azione europeo 2002 ed, infine, nel 2003 la direttiva è stata adottata: dir. 2003/98/EC, 17.11.2003, pubblicata in Gazz. Uff., 31.12.2003, L345/90.
- [17] Il cd. INSPIRE, *IN*frastruttura per *S*patial *I*nformation in *E*urope – *I*nfrastruttura per l'Informazione Territoriale in Europa – è un progetto della Commissione Europea teso a realizzare una rete di infrastrutture nazionali interoperabili per la condivisione di dati territoriali nella Comunità Europea.
- [18] Sul punto si veda il Rapporto sulla European Research Area, nella Comunicazione della Commissione COM(2014) 575 final, e la pubblicazione "The three Os", 2016, a cura della Commissione Europea, Direzione Generale per la Ricerca e l'Innovazione



**Gruppo di lavoro costituito  
con Decreto del Ministro  
del 27 Gennaio 2016**

**Rapporto pubblicato il  
28 Luglio 2016**

