



# Alfabetizzare ai dati nella società dei big e open data: una sfida formativa

## Data Literacy in the context of Big and Open Data: an educational challenge

---

Juliana E. Raffaghelli

Università degli Studi di Firenze  
julianaelisa.raffaghelli@unifi.it

### ABSTRACT

The central thesis of this paper is that to mine the treasure of big and open data, it is necessary to propose analytical basis to promote educational research and training of trainers on the issue of data literacy as skill required to operate in this context of innovation. The paper starts introducing the phenomenon of big and open data; hence, it characterizes the multi-level training needs identified along the several levels of education and training; then, it focuses on the definition of data literacy, further discussing some frameworks for its analysis and development. To this regard, in this last phase of conceptual elaboration, two points are discussed: the need to go beyond the “DigComp” framework for specific data literacy coverage (despite the very important background this framework provides); and the inadequacy of some existing frameworks to characterize data literacy at general education levels, not just as a specialist competence. Finally, the paper introduces a draft data literacy framework aligning with DigComp, as tool for educational research and practice. With a validity yet to be consolidated taking into consideration the prototyping approach, it tries to lay the foundations for progressive refinement by the scientific/practice in order to support a theme that in Italy, as in the international context, requires careful attention.

La tesi centrale del presente lavoro è che per appropriarsi del tesoro che racchiudono i big e open data, risulta necessario proporre una base analitica che orienti la ricerca educativa e la formazione dei formatori in materia di alfabetizzazione ai dati o data literacy, competenza necessaria per operare nei suddetti nuovi contesti. Il lavoro introduce l'analisi del fenomeno; caratterizza il fabbisogno formativo a più livelli del sistema d'istruzione e formazione; e quindi sposta il focus sulla definizione di data literacy, discutendo ulteriormente alcuni framework per la sua analisi e sviluppo. In particolare, discute due punti: la necessità di andare oltre il framework “DigComp” per la copertura specifica della data literacy (nonostante l'importantissimo punto di partenza che esso fornisce); e l'inadeguatezza di alcuni framework esistenti per caratterizzare l'alfabetizzazione ai dati dalla base del sistema dell'istruzione e della formazione, non solo come competenza specialistica. Infine, introduce uno strumento (framework) utile alla pratica e la ricerca educativa sulla data literacy. Di validità ancora da consolidare vista la modalità di elaborazione, tenta di porre le basi per un progressivo perfezionamento da parte della comunità scientifica e di pratica interessata, a supporto di un tema che in Italia come nel contesto internazionale, richiede sicura attenzione.

### KEYWORDS

Open Data, Big Data, Data Literacy, Educational Research.  
Open Data, Big Data, Alfabetizzazione ai Dati, Ricerca Educativa.

*Mi domandi quale sia la giusta misura della ricchezza?  
Primo avere il necessario, secondo quanto basta.  
Quis sit divitiarum modus quaeris?  
Primus habere quod necesse est, proximus quod sat est  
(Seneca, Lettere morali a Lucilio, Lettera 1,2)*

## Introduzione

La base dei flussi informativi come discorsi, politiche, scienza, istruzione, business e anche, per l'intero farsi della nostra quotidianità, sono una serie di unità informative che assumono più forme e che possiamo chiamare "dati". Ma parliamo di un concetto complesso, cui polisemia ci colpisce qualora tentiamo di districarla: alla precisione e l'obiettività che nella storia si è tentato di dare al concetto di "*datum*", oggi non abbiamo che l'imprecisione delle molteplici forme con cui i cosiddetti "dati" si presentano e vengono manipolati in successive forme via via più sintetiche, più vicine alla comunicazione umana. Ma sono i dati unità di verità, di fatti, di conoscenza? Qual è la distanza tra il dato e le metodologie che consentono di estrarre senso, ovvero ciò che ogni società e ogni individuo cerca incessantemente?

La storia ci insegna che "i dati" sono stati da sempre prodotti e manipolati agli effetti di fondamentare in modo "empirico" un certo messaggio più concettuale. Tuttavia, ciò che è decisamente cambiato è il modo con cui i dati vengono generati e soprattutto il modo attraverso il quale i dati possono essere esplorati. Mentre fino alla comparsa di Internet i dati dovevano essere generati e raccolti meccanicamente o manualmente, la progressiva digitalizzazione di servizi e attività nella società contemporanea ha portato alla generazione di masse di dati digitali. Si pensi per esempio alle ricerche statistiche sulla composizione socio-economica e demografica della popolazione: grandi indagini portate avanti attraverso censimenti con migliaia di operatori allenati all'intervista strutturata, seguita dalla compilazione di formulari dai quali i dati andavano estratti, caricati e predisposti all'elaborazione. L'avvento delle reti sociali e loro integrazione con servizi web di diverso tipo (per esempio la realizzazione di acquisti), oppure con l'"*internet of things*", crea automaticamente migliaia e migliaia di inputs digitalizzati che possono essere esplorati, estratti e convertiti in evidenza su comportamenti quotidiani, caratteristiche socio-demografiche, forme di partecipazione e coinvolgimento in processi sociali. Naturalmente, ciò che doveva essere prima indotto come processo di generazione di dati, a partire da questo fenomeno si converte in una sorta di "tesoro" che deve essere estratto, motivo per il quale si parla di "data mining" o miniera di dati. È questa la base del concetto "Big Data", che si arricchisce poi a partire da una serie di riflessioni: le problematiche legate alla gestione di masse di dati abnormi, basate fisicamente su più server (a volte, migliaia di server); le metodologie di estrazione e di conversione in "visualizzazioni" che risultino comprensibili e significative per utenti non esperti; l'uso nei vari settori di business per supportare servizi o attività produttive, oppure (e soprattutto) per generare innovazione. Quasi in modo immediato il movimento dei "Big Data" ha fatto emergere un forte dibattito relativo alla procedenza, l'uso e l'appropriazione dei dati: mentre era vero che compagnie come Facebook o LinkedIn generavano dati di proprietà privata aziendale, era altrettanto vero che i dati prodotti dalla scienza e la pubblica amministrazione dovevano essere resi pubblici. Infatti la potenzialità di uso di tali dati, poiché generati con denaro pubblico, doveva essere aperta alla società (Zuiderwijk & Janssen, 2014). Nasce così il movi-

mento Open Data, che non contrasta ma complementa il concetto di Big Data con un senso di democratizzazione della conoscenza proprio dei movimenti Open Access e Open Science.

Infatti i discorsi più entusiasti sulla facile disponibilità dei dati e la loro appropriazione hanno connesso tale situazione all'accesso alla conoscenza, alla trasparenza del governo e della scienza che potrebbero consentire nuove modalità di business e nuova occupazione. Ma soprattutto, si è parlato di nuove forme di partecipazione della cittadinanza, di consapevolezza, di controllo, addirittura di "hacking" di dati che potrebbero essere la base di forme di appropriazione della cosa pubblica e di resistenza al potere (Baack, 2015).

Ecco che l'utopia dell'accesso ai dati s'imbatta con un problema non nuovo nella sua configurazione, il problema tipico della società digitale, dell'era Internet e dell'"Open" in tutte le sue forme: il problema dell'abbondanza, che richiama immediatamente la competenza per navigare tale abbondanza. Il fatto che numerosi dataset siano estraibili o siano addirittura aperti e accessibili online in portali pubblici non li rende immediatamente comprensibili. Alcuni autori paragonano il problema dell'appropriazione dei dati da parte della cittadinanza al lungo dibattito sul *digital divide* (Gurstein, 2011); perché l'accesso ai dati da parte della cittadinanza si trasformi in controllo delle politiche pubbliche sarebbe necessario comprendere quali sono i problemi e gli oggetti tracciati attraverso i dati; avendo poi le capacità di sintesi per segnalare eventuali criticità (Zuiderwijk, Janssen, Choenni, Meijer, & Alibaks, 2012). E questo sarebbe il livello più basso di uso dei dati pubblici, se teniamo conto dei problemi di generazione e condivisione di dati tra ricercatori in un contesto di apertura della scienza (Janssen, Charalabidis, & Zuiderwijk, 2012).

In questo scenario, appare chiaro che ci sia un gap tra potenzialità intravviste dal fenomeno sociale e di innovazione tecnologica dei *big e open data*, e le pratiche correnti. È vero che l'evoluzione dei suddetti movimenti è stata velocemente oggetto di studio delle scienze sociali, in particolare scienze economiche (problemi dell'innovazione, modelli di business basati su *big data*, marketing), le scienze politiche (e-Government e Open Government, cittadinanza digitale e partecipazione civica attraverso l'appropriazione dell'informazione pubblica) e la sociologia della scienza (Open Science, accessibilità della conoscenza scientifica pubblica, nuove forme di scholarship nell'era digitale). Tuttavia, la configurazione del problema come problema *pedagogico*, ovvero di natura formativa, sembra essere poco esplorata. Per esempio, già nel 2010 Davies lamentava la mancanza di appropriazione di Open Data all'interno del settore educativo (Davies, 2010). Con alcune esperienze e dibattiti pionieri sugli Open Data come Open Educational Resources, Atenas, Havemann, & Priego (2015) proponevano i primi passi per pensare forme di uso educativo dei dati disponibili in rete.

Dunque, il problema sembra essere complesso e dovrebbe spronare un dibattito pedagogico che dovrebbe far convergere diversi ambiti della ricerca educativa, dalla ricerca sui profili emergenti e le competenze specialistiche necessarie per la copertura di nuove professionalità operanti in settori di business ad alta innovazione, al luogo che l'alfabetizzazione ai dati dovrebbe avere nella formazione di base scolastica e universitaria. Infatti, ciò che potrebbe reclamarsi alla ricerca educativa è il consolidamento di modelli di competenze, strumenti e metodologie per l'alfabetizzazione in dati progressiva dall'obbligo scolastico a situazioni di sviluppo professionale, alla valutazione d'impatto che i suddetti interventi potrebbero avere nei contesti di sviluppo socio-economico e culturale entro cui l'intervento formativo si rende necessario.

L'obiettivo dell'articolo è quindi quello di riflettere sul problema dell'alfabetizzazione ai dati posto dall'emergere del fenomeno dei *big* e *open* data per pensare un approccio sistematico alla pratica formativa e la ricerca educativa. A questo scopo, nei paragrafi successivi verrà introdotto il concetto di Big Data e Open Data, per meglio conoscere questo scenario di innovazione; ci occuperemo poi del fabbisogno formativo e le risposte date allo stesso, che caratterizzeremo come frammentarie ed *ad hoc*; per passare infine ad analizzare framework di competenze entro cui poter orientare la *data literacy*, nonché orientamenti alla pratica formativa e la ricerca educativa.

## 1. Big Data, Open Data: Il problema dell'abbondanza

Christine Borgman, nel suo libro del 2015 "Big Data, Little Data, No data", in seguito ad una lunga disamina delle caratteristiche e categorie entro cui i dati possono essere classificati, giunge ad una sintetica definizione: "*I dati sono la rappresentazione di osservazioni, oggetti, o altre entità usate come evidenza dei fenomeni studiati dalla ricerca accademica*". Questa definizione apre un primo ordine di problemi relativo al fatto che i dati non sempre sono generati dalla ricerca scientifica (in particolare, negli ultimi tempi e con l'avvento della digitalizzazione di processi e prodotti dell'attività umana). Il vocabolario *Treccani* definisce il dato come *ciò che è immediatamente presente alla conoscenza, prima di ogni forma di elaborazione*, e fa seguire una serie di esempi che partono dalla ricerca scientifica, per allargarsi ad una più generica considerazione della conoscenza umana: *i dati di un problema, i valori noti o presunti noti di talune grandezze, mediante i quali, sulla base delle relazioni e condizioni presupposte nell'enunciato del problema, ci si propone di determinare i valori incogniti di altre grandezze; elemento, in quanto offerto o acquisito o risultante da indagini e utilizzato a determinati scopi*.

Borgman ci ricorda inoltre che nessuna definizione potrebbe essere sufficiente, dal momento in cui i dati esistono nella misura di ciò che può essere fatto con essi, dei tipi di insight generano e delle scala di analisi con cui vanno trattati a seconda del tipo di fenomeno che voglia si intenda analizzare. Infatti il termine *datum/data*, tratto dal latino, compare presto nella letteratura scientifica, con i primi usi in latino verso il diciassettesimo secolo e un acceso dibattito sul uso singolare o plurale della parola in inglese verso il diciottesimo secolo. Quanto si desume da un dibattito lungo secoli è che i dati non sono una verità o una realtà: che si tratti di fatti, di fonti di evidenza, di principi alla base di un'argomentazione, si possono considerare forme di rappresentazione della *verità* o la *realtà che ci si prefigge di studiare*. Il corsivo va a sottolineare l'arbitrarietà della definizione di un dato come evidenza *empirica* di un flusso di attività umana, in particolare, di ricerca scientifica ma non solo, come vedremo più avanti. Il fatto che i dati siano normalmente associati alla ricerca scientifica si basa sul concetto stesso di metodo scientifico, attraverso il quale si tenta di tendere una linea che va dalla definizione teorica di un fenomeno, caratterizzato in un ambito disciplinare di studio, alle procedure e strumenti necessari per raccogliere *dati*, ovvero, per rappresentare e analizzare empiricamente tale fenomeno. Questa separazione è molto articolata nella ricerca scientifica, e porta con sé una grande attenzione alla definizione dei *dati* per i vari ambiti di ricerca. Eppure, ci lascia a bocca asciutta quando si tratta di una buona definizione generica e una classificazione esaustiva. Borgman ci indica una serie di domande per pensare che cosa siano i dati: Qual è l'arco temporale attraverso il quale un dato emerge, si mantiene e

sparisce? Come fanno individui, gruppi e comunità a creare, usare o selezionare i dati? Quali sono i fattori associati alle definizioni di dato in uno specifico contesto? Quali sono le funzioni che i dati vanno a rappresentare? Come variano queste considerazioni per tipo di studio, ambito disciplinare, creatori dei dati e curatori di questi dati? Come si evolvono le nozioni di dato lungo un ambito scientifico? E come stanno variando le nozioni di dato in base all'uso di tecnologie digitali per reperirli, trattarli e curarli? Tutte queste domande ci orientano a capire che i dati non sono oggetti naturali, che esistono *per sé*. I dati, conclude la studiosa, *possono essere più facilmente definiti attraverso esempi variegati*, includendo fatti, numeri, lettere, simboli. Quindi, sono solo le definizioni operative che mettono il dato in contesto e ne fanno emergere un senso trasversale per gli aspetti che tentiamo di ancorare attraverso una definizione.

Per esempio una definizione proveniente dall'ambito scientifico della ricerca spaziale indica che i dati sono *rappresentazioni reinterpretabili di informazione, formalizzate in modo tale da consentire la comunicazione, l'interpretazione o l'elaborazione*. Esempi di dati sono i bits, le tavole di numeri, i caratteri in una pagina, i suoni registrati a partire da una voce parlante, o un campione di roccia lunare (Consultative Committee for Space Data Systems, 2012). Il progetto *Data Documentation Initiative* (DDI) promuove una serie di standard per la generazione di metadati, ovvero i dati che descrivono i dati, che dovrebbero consentire una migliore gestione del ciclo di vita del dato (Data Documentation Initiative, 2012). L'approccio DDI è ampiamente usato, in particolare nell'ambito delle scienze sociali, tuttavia non fornisce una definizione di che cosa siano i dati!

Il nostro sforzo di definizione a partire dai casi particolari ci porta ora a pensare una serie di categorie che caratterizzano il dato: i gradi di elaborazione, gli origini e la preservazione, l'appartenenza a collezioni.

Per cominciare con la prima categoria, poniamo l'esempio dell'Earth Observing System – Data Information System (EOS DIS) della NASA, raggruppa i dati a seconda di come vengono elaborati, ovvero, attraverso diversi livelli di elaborazione. Tali livelli dimostrano il tipo di trattamento applicato al dato "crudo", così come rilevato, e le distinzioni sono tali da generare un livello 0 con dati già ripuliti dalla rilevazione essenziale realizzata con strumenti di osservazione astronomica. Il primo livello aggiunge tutta una serie di metadati di riferimento a intervalli di osservazione, parametri di misurazione degli strumenti, ecc. I livelli 2, 3 e 4 continuano ad aggiungere metadati per allineare i dati elaborati con griglie di analisi spazio-temporale, aggregando i dati verso la costruzione di modelli di interpretazione. Questa complessa struttura dipende dagli strumenti adottati quanto gli usi che si fanno dei dati. La maggior parte degli scienziati in ambito astrofisico utilizzano i dati del livello 4 in modo tale da poter confrontare modelli. Tuttavia, alcuni scienziati preferiscono i dati al livello 0 per operare singolarmente i vari livelli di pulizia, calibrazione, analisi degli *outliers* e riorganizzazione dei dati. Questi livelli di elaborazione implicano diverse manovre di gestione e manutenzione, ovvero di cura dei dati, che possono determinare una futura disponibilità, e dove si rende necessario mantenere molto bene i metadati e relativi algoritmi di analisi ed elaborazione. In ambiti scientifici come la fisica, la chimica e la biologia, i dati "crudi", appena raccolti, sono troppo voluminosi per poterli mantenere; di conseguenza, vi è un lavoro continuo di allineamento e revisione delle metodologie di raccolta, calibrazione di software di raccolta dati, controllo di errori di elaborazione. Questo approccio identifica bene alcuni dei problemi negli ambiti disciplinari delle scienze naturali; tuttavia non spiega né orienta le caratteristiche e difficoltà incontrate nei settori delle scienze sociali e l'ambito umanistico.

Inoltre, l'origine e i metodi di preservazione dei dati hanno dato luogo ad una classificazione, quella proposta dal US National Science Board (2005) che basandosi sugli ambiti scientifici intende raggiungere tutti gli ambiti disciplinari. Le categorie create dalla NSB sono: osservazioni (a partire dal riconoscimento, l'annotazione oppure il registro di fatti e occorrenze di fenomeni e sulla base di diversi strumenti); dati informatizzati (il prodotto di operazioni digitali *machine-readable*, ovvero codice interpretabile per una macchina); dati sperimentali (risultato di procedure interamente o parzialmente controllate, miranti a mettere alla prova un'ipotesi più o meno strutturata su fenomeni e relativi modelli teorici che li inquadrano e spiegano).

Molto rimane al di fuori di questa caratterizzazione. Infatti, la Borgman ci parla di una quarta categoria di dati, i registri o *records*, che include forme di dati non classificabili nelle prime tre categorie, come documentazione pubblica, registri di attività economica, pubblica, privata; libri e testi di ogni genere; materiale di archivio; audio o video; opere artistiche; reperti archeologici, ecc.

Naturalmente i limiti tra questi dati sono porosi e lo diventeranno sempre di più: al giorno di oggi la quantità di dati che emergono da procedure informatizzate, implica una massa di dati digitali, non necessariamente *machine-readable*, ma potenzialmente collocabili in questa categoria. Una volta definiti questi ambiti di produzione di dati, la questione aperta sono le infrastrutture e metodi di cura di questi dati.

L'orientamento è chiaramente a rendere tutti questi dati navigabili e fruibili dal settore pubblico e privato nel proprio sforzo innovativo (Kitchin, 2014). È in questo senso che compare un dibattito crescente al quale si orienta attraverso il termine *big data*. Questo termine pare essere stato coniato già negli anni '90 (Lohr, 2013) ma è a partire dai 2000 che acquisisce senso. Se la scala di produzione di dati fino al 2003, per tutta la storia dell'umanità era stata di 5 exabyte, da quel punto in poi questo volume informativo viene creato ogni due anni (Hal Varian, economista di Google, citato da Smolan & Erwitte (2012)). Una definizione ampiamente adottata è quella del Gartner IT dictionary delle 3V: Alto volume, alta velocità, alta varietà di informazioni compongono le masse digitali (Gartner Inc., 2013). Per quanto riguarda il volume di dati, non si parla di "campionamenti" di un dato universo, ma di continuo flusso e *tracking* di fenomeni. Si pensi per esempio ai dati emergenti da Twitter. Relativamente alla velocità, possiamo dire che i dati possono essere estratti in modo immediato: continuando con l'esempio di Twitter, esistono una varietà di applicativi web che consentono di estrarre i dati di profilo, network, tipologia di links tra membri di una rete, in tempo reale. Infine, per quanto riguarda la varietà, i *big data* si possono presentare come testo, video, audio, e potrebbero comporre nuove unità di dati attraverso la *data fusion*. Sempre in base all'esempio di Twitter, osserviamo che i tweets possono essere multimodali (testo/audio/video/immagine) e che le analisi potrebbero basarsi sull'estrazione sia di testo, sia di numeri, tags, oppure video e immagini. A questa definizione, si sono poi aggiunti due elementi, ovvero: la generazione di *dataset* (matrici di dati analizzabili) leggibili da programmi informatici ovvero *machine-readable*; e la produzione gratuita del dato a partire da un servizio esistente che crea le basi per la continua generazione di dati, ovvero il *digital footprint*. Un esempio in questo senso lo compongono le *wearable technologies* e gli applicativi per *smartphone* con geolocalizzazione, che consentono la raccolta di dati relativi alla mobilità delle persone, a parametri fisiologici (per esempio ritmo cardiaco e respiratorio nella corsa), a tipologia di attività sociale (condivisione di fotografie lungo un percorso tracciato di corsa). I processi innovativi legati ai *big data* hanno richiesto particolarmente l'esplorazione delle architetture e la quali-

tà di dati generati, per poi focalizzare le metodologie di estrazione di dati come *machine learning* e *natural language processing*, a supporto di forme di business intelligenti basati sulla continua predizione di comportamenti e interessi degli utenti a partire dai dati da loro stessi generati. Una forma molto frequente sono i *recommender system*, che basandosi sulle suddette forme di analisi fanno comparire raccomandazioni di servizi o pubblicità nelle pagine web visitate da un utente (in particolare social network). La visualizzazione delle analisi condotte a partire dalle masse di dati tracciate è stato un altro elemento di attenzione per le comunità scientifiche e professionali in quanto l'aspetto più importante è comunicare un messaggio di rilevanza all'utente finale. Che si tratti di forme di *marketing* dove il dato viene trasformato in un avviso pubblicitario; o di rappresentazioni avanzate delle proprie reti professionali su LinkedIn Labs, la visualizzazione genera un nuovo livello informativo dove i dati estratti diventano grafici, reti, raccomandazioni, mappe, infografie. Le forme più avanzate sono quelle legate a interfacce che consentono di realizzare *query* sui dati "in vivo" e quindi portare ad una visualizzazione personalizzata a seconda dei parametri fissati dall'utente (EMC Education Services, 2015). Come lo indica la Comunicazione del Parlamento Europeo che pubblica la strategia sui Big Data per incentivare la transizione verso un'economia *data-driven* [COM(2014)442 final] "*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*".

Tuttavia, tutta questa ricchezza di dati potrebbe non essere sempre disponibile a tutti, si pensi solo ai dati tracciati e curati a partire dai servizi web come quelli dei giganti Google, Youtube, Facebook, LinkedIn etc. cui sedi non sono localizzabili in Europa.

Con la filosofia di accesso alla conoscenza, del movimento Open Source e Open Access, che accompagna poi la nascita dell'Open Education e soprattutto della Open Science (Raffaghelli, 2014), nasce poi il movimento Open Data. Secondo la Open Knowledge Foundation, l'apertura della conoscenza risiede in tre elementi fondamentali, ovvero la disponibilità di accesso, il riutilizzo e la redistribuzione e la partecipazione universale alla generazione di conoscenza a partire da altre risorse liberamente trovate in rete (OKF, n.d.). Tali dimensioni si applicano poi ai dati; se consideriamo la discussione iniziale sulla natura dei dati, riusciamo ad immaginare gli open data come unità informative alla base di attività di diverso tipo, che possano essere liberamente redistribuiti e condivisi, ma soprattutto riutilizzati per la ricerca e l'innovazione. Tale è lo spirito della definizione di Open Data che ci presenta ancora la OKF (2014) in un manuale orientativo che introduce casi, soluzioni e aree aperte al continuo sviluppo della comunità attiva per l'apertura dei dati. Infatti, in questo manuale si enfatizza la necessità di produrre dati che siano accessibili e senza limitazioni basate sull'identità e sull'intenzione dell'utente; che vengano presentati in formato grezzo e digitale e *machine readable*, ovvero in formati che consentano l'interoperabilità con altri dati (per esempio file JSON, CSV, TXT); che siano liberi da restrizioni sull'uso o sulla redistribuzione, nelle loro condizioni di licenza. Infatti, la questione delle licenze non è minore, poiché il rilascio senza licenze potrebbe bloccare o addirittura vietare il riutilizzo dei dati anche ai fini commerciali. Oltre l'uso delle licenze Creative Commons (di preferenza CC-BY-SA), alcuni Stati hanno sviluppato licenze specifiche per la pubblicazione dei dati. In Italia, è stata proposta a tale proposito la Italian Open Data Licence (IODL).

Un tale attivismo viene supportato da due ambiti: uno quello della ricerca scientifica e le nuove forme di democratizzazione della stessa, ovvero l'Open

Science; e il secondo da parte della governanza aperta o Open Government. Entrambi gli ambiti hanno un aspetto in comune: l'uso di risorse pubblici per produrre beni e servizi ad altissimo impatto sociale e di un altrettanto alto costo: la premessa del movimento Open Data si basa sul fatto che tale costo deve essere restituito ad una cittadinanza attiva e coinvolta nell'amministrazione della cosa pubblica.

Rispetto all'Open Science, il trend di digitalizzazione della scienza non è nuovo (si parla di eScience già dai primi 2000), con l'enfasi posta nella velocità di condivisione di informazioni miranti alla collaborazione della comunità scientifica internazionale. Ma la possibilità di disporre i dati nel cloud, in modo riutilizzabile anonimamente poiché aperti, è un fenomeno più recente cui importanza crescente viene rispecchiata nell'enfasi posta dalla Commissione Europea per formulare politiche di apertura di dati. Nel 2015 si realizza una consultazione pubblica "Science 2.0, science in transition" (European Commission, 2015), nel contesto dell'Agenda Digitale Europea che diviene poi il fulcro del nuovo documento di orientamento delle politiche di comunicazione e condivisione della conoscenza scientifica "Open Innovation, Open Science, Open to the World" (European Commission, 2016a). Inoltre, gli orientamenti si traducono in framework per promuovere osservatori di apertura della scienza: in particolare, si fa riferimento all'Open Science Monitor (European Commission, 2016b). All'interno di questa struttura, i dati aperti acquisiscono cruciale importanza, e si promuove così il deposito di *dataset* provenienti dalla ricerca scientifica sia in *IR* (repository istituzionali, solitamente gestiti dalle biblioteche) che in portali per la condivisione di dati come *Zenodo*, *GitHub* oppure *Figshare*. L'adozione libera da parte dei ricercatori delle cosiddette *Academic Social Networks* come *ResearchGate* e *Academia.edu* rimane un aspetto controverso (poiché sebbene il profilo e le risorse in esso caricate appartengono al ricercatore, i dati (grezzi e dataset) possono essere gestiti dalle ASNS a scopo del proprio modello di business (Manca & Raffaghelli, 2017).

Il concetto di Open Government si lega invece alla ormai lunga storia di apertura del governo pubblico, la partecipazione della cittadinanza e l'empowerment. Se la trasparenza della gestione pubblica era già discussa con i primi '90, la digitalizzazione della pubblica amministrazione ha portato al reclamo di una completa trasparenza anche a partire dalle unità informative da essa prodotte quotidianamente: i dati (Davies, 2010). I concetti chiave del governo aperto sono la partecipazione, la trasparenza e la collaborazione; e i suoi strumenti, giustamente, i dati aperti e l'uso di reti sociali e altri ambienti web per la comunicazione diretta con la cittadinanza (Lathrop & Ruma, 2010). Da metà dei 2000 in poi, in diverse realtà europee si consolidano lentamente portali per l'eGovernment che poi lanciano portali per i dati aperti. L'Unione Europea lancia il proprio portale per i dati aperti<sup>1</sup> nel 2011. Questo portale, al momento attuale, supporta tutta la catena di valore associata ai dati aperti: dalla pubblicazione dei dati all'accesso aperto al riutilizzo degli stessi. Negli ultimi anni si è posta estrema attenzione alla classificazione dei formati digitali e la struttura di metadati associati ai dataset, che nel portale europeo conta con un Thesaurus (Eurovoc<sup>2</sup>). I dataset, che ad oggi sono 754.522, e<sup>3</sup> sono navigabili per settore di attività (governo, finanza, busi-

1 <https://www.europeandataportal.eu/>, Settembre 2017

2 <http://eurovoc.europa.eu/>, Settembre 2017

3 Lettura del 25 Settembre 2017.

ness, istruzione e formazione, ambiente, agricoltura, trasporto e mobilità, ecc.) e per Stati Membro dell'UE. Inoltre, esiste una casistica molto ricca di tipologie di uso del portale che dovrebbe portare all'ulteriore generazione di servizi e prodotti. Per quanto riguarda il caso italiano, il portale dei dati aperti si apriva, anch'esso nel 2011<sup>4</sup>. Tuttavia, il portale resta più legato alla prima fase di valorizzazione dei dati prodotti che alle connessioni con esempi di uso. Ad ogni modo il portale propone uno spazio di connessione con i dati nazionali, regionali e comunali, che via via si arricchiscono. Il portale raggruppa ad oggi 13602 dataset<sup>5</sup>.

Nonostante l'enorme progresso nelle modalità di condivisione e le politiche e pratiche di riutilizzo, il portale *Open Data Barometer*<sup>6</sup> sembra denunciare una situazione di disequilibri. Questo portale ci consente di analizzare in modo dinamico lo stato di produzione di dati aperti, considerando alcune dimensioni di implementazioni delle politiche di apertura, e l'impatto sulla società (in termini di uso dei dati per la generazione di servizi o prodotti).

La figura 1 introduce un estratto del 4to rapporto dell'Open Data Barometer. Ci sono due elementi da considerare: in alto, un grafico correlazionale, e in basso, un ranking. Il grafico (primo elemento) ci mostra la correlazione tra il coinvolgimento della società civile con il movimento dei dati aperti e contro l'impatto (ovvero l'uso ai fini innovativi). Le due variabili hanno una correlazione positiva con una certa dispersione iniziale, tuttavia si può affermare che a maggiore coinvolgimento, maggiore innovazione basata su Open Data. Il ranking sottostante (secondo elemento) ci consente di osservare i primi tre "best performer" in materia di politiche di apertura di dati (Gran Bretagna, Canada, Francia); a confronto di "good performer" tra cui troviamo la Svizzera (19 posto) e l'Italia (20). Qual è la differenza tra questi Stati? Si osserva una minore predisposizione e una minore implementazione della pubblicazione di dati aperti (si vedano per esempio i 100 punti di implementazione del Regno Unito contro i 51 dell'Italia) che si associa ulteriormente ad un minore impatto (94 punti contro 37 punti degli stati menzionati).



Figura 1 – Composizione di risultati basata sull'Open Data Barometer 4th report

- 4 <http://www.dati.gov.it/>, Settembre 2017
- 5 Lettura del 25 settembre 2017
- 6 <http://opendatabarometer.org/>

Quest'informazione ci consente di pensare che esista un fabbisogno formativo, il quale emerge dinanzi alla necessità di portare le pratiche a livelli migliori. Ma come si sta rispondendo a un tale problema? Nella sezione successiva ci occuperemo di tracciare, sebbene in modo non esaustivo, alcune esperienze e approcci che stanno tentando di muoversi nella direzione di fornire *skills* per il processo innovativo promosso dai *Big e Open Data*.

## 2. Il fabbisogno formativo

Il mondo dell'istruzione e la formazione ha reagito quasi immediatamente all'avvento del fenomeno dei *big e open data*, con lo *skills' gap* da esso generato. Tuttavia, il movimento è stato più reattivo che proattivo e la domanda che la comunità della ricerca educativa dovrebbe porsi è: quale strategia implementare a tutti i livelli di istruzione e verso un'ottica *lifelong learning*? Ci concentreremo prima su alcuni casi emergenti, che ci consentiranno di osservare la ancor frammentaria risposta del sistema dell'istruzione e della formazione in questo senso. In particolare, ci soffermeremo sulla nuova figura professionale del *data scientist*, legata alla gestione di dati e le prime esperienze formative tese a formare la suddetta professionalità; le problematiche di formazione all'uso degli *open data* nella ricerca scientifica; le esperienze informali di attivismo sul *open government* che portano alla comprensione del fenomeno e la sua appropriazione da parte della cittadinanza. Discuteremo infine le prime politiche di inquadramento di pratiche nell'ambito formale dell'istruzione per l'inclusione dell'alfabetizzazione di dati.

*Data scientist, una professionalità del futuro.* Nel 2014, Udacity pubblicava un post sul proprio blog intitolato "8 skill per necessarie per essere un *data scientist*"<sup>7</sup>. Tra queste si elencavano:

- Gli *strumenti di base* che implicano la programmazione statistica su linguaggi come R oppure Python, e la conoscenza di linguaggi per l'interrogazione di database come SQL;
- le *statistiche base*, per la comprensione e uso di nozioni come probabilità, p-value, statistica inferenziale, distribuzioni, ecc.;
- la conoscenza del concetto e tecniche di *machine learning*, per comprendere quando applicarle nella navigazione della complessità attraverso algoritmi che consentano di anticipare, riprodurre, raggruppare, ristrutturare fenomeni come parte di programmi che "apprendono";
- *L'algebra lineare e il calcolo multilineare* alla base della statistica e la programmazione. La comprensione di questi concetti sarebbe più importante in settori di business dove il prodotto è definito dai dati e da piccoli migliorie che consentono la predittività delle performance di un algoritmo ottenendo comunque ingenti benefici.
- Il "*data munging*", ovvero la pulizia di dati in stato iniziale molto caotico che includono valori iniziali mancanti, stringhe o formati incoerenti.
- La *visualizzazione e la comunicazione basata sui dati*, aspetto incredibilmente importante per descrivere risultati basati su dati verso utenti tecnici e non tecnici.

7 <https://blog.udacity.com/2014/11/data-science-job-skills.html>, Settembre 2017

- La settima e penultima skill è quella di una adeguata conoscenza di *software engineering*, come background generale per lo sviluppo di prodotti data-driven.
- Infine, ma non meno importante, l'abilità di pensare come data scientist, in quanto "*data-driven problem solver*": qualcuno che sa individuare problemi rilevanti per la compagnia, e ricondurli allo sviluppo di soluzioni basate sui dati.

Queste skill, informalmente individuate da un *MOOC provider* mostravano già una tendenza importante: quella di offrire un'offerta formativa internazionale, massificata e per tale motivo funzionale al modello di business MOOC, che indicava un rapporto stretto con il mercato del lavoro. Infatti, altri noti *MOOC provider* come Coursera, edX, FutureLearn. Coursera<sup>8</sup> offre alla data odierna 5 specializzazioni (percorsi composti da 5 a 10 MOOC) su materie che formano le skills e conoscenze sopra elencate in *Data Science*, e un totale di 961 corsi che rientrano in questa categoria. edX<sup>9</sup>, a suo tempo, offre 281 corsi relativi alla materia, con 12 "Micro-masters" che sono composti da MOOC e attività progettuali, nonché 7 "Professional Certificates" ovvero percorsi brevi basati su competenze richieste dalle aziende. FutureLearn<sup>10</sup>, il MOOC provider britannico, offre 42 corsi; Miriada<sup>11</sup>, il provider spagnolo, 10 corsi; Eduopen<sup>12</sup>, provider italiano, 5 corsi. Essendo i MOOC un'offerta formativa fluida, la risposta ad un mercato del lavoro dinamico è stata evidentemente rapida e immediatamente accolta. Tuttavia, tale offerta non raggiunge i percorsi universitari in modo immediato, e solo negli ultimi tre anni si verifica la progressiva apertura di Master universitari e Insegnamenti all'interno dei Corsi di Laurea che rispecchiano l'approccio interdisciplinare e basato su problemi richiesto dalla *data science*. Il Rapporto Big Data del MIUR (2016) elenca per lo meno 4 università con CdL triennale e magistrale contenente la tematica dei *big data*, 4 università con Master di I livello relativi, e 7 università con Master di II livello. Indica poi l'esistenza di 4 Lauree Magistrali complete e 11 curricula specifici per Big Data, 5 LM complete e 7 curricula su Cybersecurity, nonché la presenza nei programmi (97 insegnamenti per Big Data, più di 200 sulla sicurezza informatica). Come possiamo osservare, la diffusione è al livello meno strutturato dell'alta formazione (corsi di Master, curricula e insegnamenti all'interno di CdL), orientandosi a competenze altamente specializzate. Infatti, come lo indica il Rapporto (p. 28)

*"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"*

Questa situazione sta mettendo in evidenza la necessità di considerare il problema formativo a più livelli e da più prospettive, per attivare forme interdisciplinari, legate a problemi sociali e culturali, inseriti in più ambiti produttivi e dei ser-

8 <https://www.coursera.org/>, Settembre 2017

9 <https://www.edx.org/>, Settembre 2017

10 <https://www.futurelearn.com/>, Settembre 2017

11 <https://miriadax.net/>, Settembre 2017

12 <https://learn.eduopen.org/>, Settembre 2017

vizi. Si pensi per esempio alle skills trasversali delle persone che lavoreranno nel campo delle *learning analytics*. A monte dei processi di mineria di dati, di analisi statistiche e di visualizzazioni, vi è un problema educativo che richiede la adeguata caratterizzazione dei costrutti (per esempio, abbandono o successo formativo).

*La formazione agli Open Data in ambito scientifico.* Un altro settore che è cresciuto in modo poco strutturato è stato quello che orienta alla formazione di competenze digitali specialistiche per la ricerca (Raffaghelli, 2017). Dentro un panorama di promozione della ricerca aperta e la pubblicazione *Open Access*, diverse iniziative del Programma quadro per la ricerca europea Horizon2020 hanno orientato alla generazione di azioni e interventi per la formazione dei ricercatori. Per esempio il concept paper “Digital Science at Horizon 2020” della DG Connect (Commissione Europea, 2013) offre una visione di scienza digitale, un framework per lo sviluppo di conoscenze e competenze e un numero di dimensioni operative che dovrebbero guidare i ricercatori nella propria pratica professionale di scienza aperta; in questo contesto, i dati sono menzionati. Invece la comunicazione della Commissione Europea (COM 2016 178 final) sulle nuove infrastrutture europee al supporto della scienza sottolineano la necessità di attivarsi per l’implementazione dell’iniziativa European Cloud, con lo scopo di generare infrastrutture adeguate per la condivisione di dati, la loro gestione e la formazione, alfabetizzazione e rafforzamento di skills nella materia (p.6). Più recentemente è stato creato un Gruppo di Esperti di Alto Livello sull’European Science Cloud, il quale ha subito raccomandato la necessità di training sulla scienza aperta (Ayrís et al., 2016). Inoltre, in Marzo 2016 si apriva lo spazio Futurium della DG Connect per la consultazione pubblica e il dibattito sulla cosiddetta “Open Scholarship”; si diceva in questo documento: *...cooperate with the New Skills and Professions group to design an action plan for training a new generation of scholars and shaping model policies for career development in Open Scholarship* (Matt, 2016). Infine, ad Ottobre 2015 il programma di lavoro sulla Scienza con e per la Società includeva una call per finanziare progetti miranti al supporto della formazione dei ricercatori per l’Open Science (European Commission Decision, 2016), che si chiudeva ad Ottobre 2016. All’interno di questo contesto, come nel caso dei *Big Data*, diversi portali per la promozione degli *Open Data* hanno lanciato modalità flessibili e maggiormente digitalizzate per introdurre le problematiche di generazione, gestione, condivisione e riutilizzo di dati aperti. Si veda per esempio i casi del portale EUDAT<sup>13</sup> che promuove attività in autoformazione così come webinar specifici; oppure l’area Open Data offerta dal progetto FOSTER<sup>14</sup>. Non mancava l’offerta formativa basata su MOOC (per esempio il corso “Research Data Management and Sharing” di Coursera<sup>15</sup>). In tutti i casi si osserva uno sguardo frammentario, talvolta più centrato sul flusso informativo e la pubblicazione dei dataset scientifici come prodotti della comunicazione scientifica; talvolta sulla problematica etica di condivisione di dati (particolarmente inerenti alle scienze umanistico-sociali) o la gestione nei casi di dataset distribuiti (per esempio nella ricerca spaziale). Quanto emerge da questa superficiale mappatura è un’ingente necessità di pensare invece la formazione di base dei ricercatori (a partire dal terzo ciclo Bologna, ovvero i dottorati) e di rafforzarla lungo l’intero arco di vita professionale dei ricercatori, come parte di un approccio al-

13 <https://www.eudat.eu/>, Settembre 2017

14 <https://www.fosteropenscience.eu/>, Settembre 2017

15 <https://www.coursera.org/learn/data-management>, Settembre 2017

la *digital scholarship*.

Tra il formale e l'informale: l'attivismo per l'apertura di dati. In particolare la formazione agli Open Data si è cimentata su interventi che non nascono come attività formative, ma cui obiettivi implicano necessariamente forme di apprendimento soprattutto informale e non formale. Ci riferiremo in particolare all'attivismo per l'apertura di dati, che si presenta nelle forme di *Hackaton*, *Data Expeditions*, *Contest* e altri eventi per promuovere il coinvolgimento della cittadinanza con forme di Open Government e di scienza partecipata e responsabile. In particolare, l'obiettivo di questi eventi sarebbe supportare forme avanzate di navigazione in portali di dati aperti; ma le metodologie si diramano a seconda che si tratti di competenze più legate alle forme di estrazione e miniera e l'ulteriore sviluppo informatico e statistico (*Hackaton* e *Contest*), alla navigazione, consulta, comprensione e analisi semplice per il controllo dello stato di avanzamento di opere pubbliche (*Data Expeditions*). Ci limiteremo a menzionare alcuni significativi esempi che possano rendere l'idea del *modus operandi*, per riflettere quindi sugli impatti formativi.

Per quanto riguarda il primo tipo di interventi, un buon esempio è il caso delle attività promosse dalla Open Knowledge Foundation Italia a partire dalla Fondazione Bruno Kessler. Questo gruppo ha promosso diversi "Open Data Camp" e "Hackaton" con studenti della scuola secondaria e universitari, insieme ad aziende e pubblica amministrazione. La struttura degli eventi è competitiva e può riguardare problemi autentici e immediati del territorio in cui ci si trova ad inserire l'evento. Per esempio, nel caso dell'evento del "Gran Sasso" all'Aquila (2017)<sup>16</sup>, il motivo è stato la ricerca di soluzioni per i terremotati, con due linee di lavoro o *track*. La prima "*empowering APPs*", relativa allo sviluppo di crowdmapping (mappe cui dati vengono aggiornati dagli utenti attraverso diversi meccanismi quali le reti sociali), monitoraggio civico, servizi per il rilancio della città e nuove routine per l'uso di dati, da inserire nella piattaforma Open Data Ricostruzione. La seconda "*revealing images and analysis*" mirante a creare infografiche, arte digitale *data-driven* e analisi che supportino la comprensione dello stato di ricostruzione della città e potenzialità di sviluppo futuro. Altre iniziative sono *Open Data Day* (per le quali ci sono finanziamenti offerti dalla Open Knowledge) gli *Sport Hackdays*. La continuità di questi eventi dal 2012 (ormai 5 anni), orienta a capirne la sua evoluzione ed importanza, sebbene non si diano particolari indicazioni di impatto a partire da questi eventi (numero di partecipanti, tipologia di servizi sviluppati). Inoltre, questi eventi ricalcano lo sviluppo internazionale di iniziative che adottano simili metodologie. Quanto emerge dalla letteratura esistente sul tema, ci orienta a capire che con alcune criticità legate alla generazione di circoli ristretti ed elitari di "culture giovanili" (lo stereotipo del *hacker*), nonché una certa entropia di questi eventi (più energie nell'immagine e la promozione che effettivi risultati) l'apprendimento informale legato all'innovazione aperta è reale ed ha potenti e immediati effetti sulle persone coinvolte, che potrebbero percorrere traiettorie a lungo termine che le porterebbe sì all'acquisizione di competenze e la generazione di innovazione (Johnson & Robinson, 2014).

In Italia, un'altro caso di estremo interesse è "*A Scuola di Open Coesione*" [ASOC], si autodefinisce come *una sfida didattica e civica rivolta a studenti e docenti di istituti secondari di secondo grado: partendo dall'analisi di informazioni*

16 Si veda per esempio <https://it.okfn.org/2017/06/23/hackathon-al-gran-sasso-science-institute-iscrivetevi-avete-tempo-fino-al-1-luglio/>, Settembre 2017.

e dati in formato aperto pubblicati sul portale OpenCoesione<sup>17</sup>, abilita gli studenti a scoprire come i fondi pubblici vengono spesi sul proprio territorio e a coinvolgere la cittadinanza nella verifica e nella discussione della loro efficacia<sup>18</sup>. Chiaramente orientata all'attivismo civico attraverso i dati aperti, l'interesse di quest'iniziativa è il suo formato didattico che implica la progettualità a scuola attraverso compiti autentici dove gli Open Data diventano risorse educative aperte (Open Educational Resources). L'impianto è basato su un'efficace modello comunicativo sul web e una community di scuole, quindi una comunità di apprendimento professionale dove gli insegnanti interagiscono con il gruppo di esperti ASOC. Il modello si arricchisce da un'esperienza ormai pluriennale (i primi progetti sono stati lanciati nel 2015) e una continua mappatura delle scuole partecipanti, che configurano la base di conoscenza ed evidenza empirica trainante per la partecipazione di nuove scuole. Il formato fa leva, infine, sull'affiancamento di un osservatorio europeo (la rete EDIC, Centri d'informazione Europe Direct) grazie ad un accordo della rappresentanza italiana presso la Commissione Europea. Secondo la mappatura degli stessi ASOC, la rete si compone ormai da 120 scuole, 56 associazioni, 28 centri EDIC, 17 enti sul territorio<sup>19</sup>. Si tratta di una rete cui nodi si allargano a tutto il territorio nazionale italiano. Il modello d'intervento didattico si basa su 5 fasi miranti al controllo civico di attività pubbliche attraverso l'uso di Open Data. La prima fase è quella della "progettazione", nella quale gli studenti sono guidati a scoprire, discutere e capire i problemi del territorio e decidere il tema del progetto di monitoraggio civico. L'attività viene svolta naturalmente in gruppo e quindi si adottano strategie di *cooperative learning* per la distribuzione dei compiti e organizzazione del lavoro. La seconda fase "approfondire" mira a formare i ragazzi alle tecniche di *data journalism*, ovvero giornalismo basato sui dati, nonché tecniche di raccolta dati dal portale Open Coesione. Così si preparano a una *data expedition* ovvero una spedizione di raccolta dati e di monitoraggio civico sui portali della pubblica amministrazione. La terza fase "analizzare", punta al miglioramento delle skill di information literacy e il controllo delle fonti, andando a confrontare i territori e quindi a costruire indicatori specifici per il progetto di monitoraggio specifico. La quarta fase "esplorare", si basa sulla pulizia delle domande che indagano sull'indicatore scelto, nonché sull'importanza della documentazione dell'espedizione di controllo, per completare un report che diventa un vero e proprio prodotto generato dagli studenti (*learners generated content*) per la sua pubblicazione sul sito *Monithon*<sup>20</sup>. La quinta fase "raccontare", si pone l'obiettivo di sviluppare tecniche di *data storytelling*, che orientano a narrare le esperienze e preparare eventi pubblici per presentare le esperienze degli studenti. Infine, la sesta fase "coinvolgere" è quella in cui si dialoga con gli enti sul territorio raccontando i risultati dell'attività di mo-

17 <http://www.opencoessione.gov.it/>, Settembre 2017. *OpenCoesione è il portale sull'attuazione dei progetti finanziati dalle politiche di coesione in Italia. Sono navigabili dati su risorse assegnate e spese, localizzazioni, ambiti tematici, soggetti programmatori e attuatori, tempi di realizzazione e pagamenti dei singoli progetti. Tutti possono così valutare come le risorse vengono utilizzate rispetto ai bisogni dei territori.* Dalla lettura a Settembre 2017, e per la programmazione-2014-2020, risultano 943.950 progetti monitorati.

18 <http://www.ascuoladiopencoessione.it/>, Settembre 2017.

19 Rete ASOC 2015-16 <https://graphcommons.com/graphs/49f31fa2-ff20-4e59-a3c0-282ef43a09d4>, Settembre 2017.

20 <http://www.monithon.it/>, Settembre 2017.

monitoraggio pubblico, puntando a sensibilizzare la comunità sulle problematiche scoperte.

La forza dell'impianto è indubbia: si basa su un'architettura didattica *problem-based*, cui problemi sono autentici (il problema che merita il monitoraggio civico); adotta step specifici e tecniche che attivano il dialogo interdisciplinare (la conoscenza della materia monitorata, per esempio dati ambientali o socio-demografici, associata ai dati statistici elaborati sulla base dei dati aperti grezzi) e la formazione di competenze digitali per la manipolazione dei dati. Espone i ragazzi al problema etico e politico della cittadinanza attiva, quindi li forma alla partecipazione nella vita civica e così attiva forme di *agentività* nel diventare, giustamente, agenti che incidono non solo all'interno dell'aula, ma anche oltre, nel dialogo con il mondo adulto (fasi di racconto e coinvolgimento). Questi elementi sono tutti in linea con le migliori raccomandazioni del mondo pedagogico: partire da problemi autentici e rilevanti, attivare l'apprendimento collaborativo e sociale, generare prodotti autentici a partire dal processo di apprendimento. Le criticità possono essere legate alla necessità di attivare forme istituzionali che richiedono sicuramente una presenza collaborativa anche da parte del personale docente, vista la necessaria interdisciplinarietà. Inoltre, racchiude una complessità logistica (uso di tecnologie, visite di monitoraggio) che potrebbe sfuggire al singolo docente.

Come abbiamo osservato l'attivismo per l'apertura dei dati e il monitoraggio civico apre a forme estremamente dinamiche di apprendimento di skills collegate all'alfabetizzazione dei dati. Auspicabilmente, esperienze come quella di ASOC stanno già sponendo dibattiti di tipo metodologico e didattico. Tuttavia, va ricordato che i promotori di questi eventi e attività non hanno come primissimo scopo la ricerca didattica ed educativa, quindi molto di questo potenziale resta per lo meno inosservato e disconnesso dalla valutazione d'impatto formativo.

### 3. Il vero problema: una formazione di base per la "data literacy"

Le varie formative iniziative fin qui analizzate ci mostrano un panorama di intensa attività che cerca di dare risposta a diversi obiettivi di innovazione produttiva, organizzativa e politico sociale. Tuttavia, un approccio pedagogico sistematico che riesca a coprire i gap formativi in modo programmatico e che tenda ad unire i vari tasselli dell'attuale *puzzle*, richiede un ulteriore sforzo di elaborazione concettuale. Non partiamo da una mancata attenzione alla questione, ma da una frammentarietà delle linee di azione e pensiero. Il primo elemento da cogliere è quello della verifica, già menzionata, del gruppo "Big Data" del MIUR che nel proprio rapporto del 28 Luglio 2016 indicano già la necessità di:

- Inserire la formazione al data science a livello universitario (CdS, LM, Master I e II livello).
- Attenzione a Big Data nei progetti di ricerca supportati da programmi nazionali.
- Maggiore attenzione a livello della Scuola alla *data literacy* nel contesto degli approcci per la competenza digitale.
- Miglioramento della diffusione del patrimonio informativo del MIUR come base per la governance e sviluppo del sistema scolastico e la ricerca educativa.

Queste indicazioni orientano pratiche innovative a livello di programmazione dell'istruzione formale a tutti i livelli. Inoltre al interno del Programma Nazionale per la Scuola Digitale, a Settembre 2016 si lancia il Bando sui curricula digi-

tali che include come tema chiave *Big e Open Data*, come problema che richiede la progettualità innovativa in ambito didattico.

Tuttavia, come si può facilmente evincere, questo tipo di approccio sta chiamando ad un'azione dal basso che poi genera una ricorsività progettuale che può o meno orientare approcci ragionati e scalabili. Come lo testimonia ampiamente il settore delle tecnologie educative, le *call for proposals* spesso creano movimenti di entusiasti e avanguardisti, e meno frequentemente, *mainstreaming* ovvero azioni semplici che però portano gli investimenti a sistema.

La discussione deve necessariamente spostarsi sugli strumenti di analisi e promozione della formazione di competenze di base e avanzate in relazione ad una alfabetizzazione in dati o *data literacy*, legate anche alle epistemologie sottostanti alle forme di conoscenza non più basate sulla "raccolta", bensì sull'estrazione dei dati. Quest'ultimo approccio, implica nuove forme di valutazione e di comprensione dei limiti del dato come elemento ontologico, come entità che nel configurarsi assume aspetti concettuali, ideologici, e addirittura di potere.

Ma che cos'è la *data literacy*? Nel rapporto canadese del 2015 promosso dal Consiglio Nazionale delle Scienze Umanistiche e Sociali del 2015 sulla *data literacy* questa viene come l'abilità di raccogliere, gestire, valutare e usare i dati in contesti applicativi, in modo critico (Ridsdale et al., 2015). Tale abilità diventa competenza quando può essere messa in contesto sia a livello accademico (la manipolazione ai fini della comunicazione accademica), all'interno del proprio settore disciplinare; sia a livello professionale, quando i dati vengono utilizzati per informare processi e decisioni (livello più essenziale) o per generare prodotti e servizi (livello avanzato e di innovazione). Le diverse definizioni esistenti coincidono sugli elementi chiave: estrazione, gestione ed elaborazione, approccio etico critico. Tradizionalmente, il costrutto è rimasto implicito nel più ampio concetto di *information literacy*, e cresciuto all'interno dello stesso fino al momento di ottenere uno status proprio. Per esempio, i modelli di *information literacy* mirano a coltivare: a) abilità di ricerca e analisi informative generiche; b) uso critico dell'informazione all'interno di settings diversificati; c) uso critico ed etico dell'informazione per trasformare il mondo (Lupton & Bruce, 2010). Movendo da queste grandi categorie, gli approcci all'alfabetizzazione sui dati si sono focalizzati inizialmente in attività da tenersi in contesti formali, particolarmente legate all'insegnamento delle scienze e la matematica (Stephenson & Schifter Caravello, 2007). Tuttavia, il dibattito nel contesto dell'*information literacy* ha posto le basi per suggerire approcci più ricchi, oltre l'apprendimento formale, considerando non solo aspetti di tipo cognitivo ma anche etico e creativo (Carlson, Fosmire, Miller, & Nelson, 2011). Secondo la revisione della letteratura di Maybee & Zilinski (2015) sulla base dell'analisi di 8 framework per la *data literacy* ha individuato i seguenti elementi:

1. Consapevolezza: comprendere che cosa sono i dati e quale sia il loro ruolo per la società.
2. Accesso: Comprendere come identificare, localizzare e usare appropriatamente *datasets* e *databases*.<sup>21</sup>

21 Un **dataset** (o **data set**) è una collezione di dati. Più comunemente un dataset costituisce un insieme di dati strutturati in forma relazionale, cioè corrisponde al contenuto di una singola tabella di un **database**, oppure ad una singola matrice di dati statistici, in cui ogni colonna della tabella rappresenta una particolare variabile, e ogni riga corri-

3. Coinvolgimento: Valutare, analizzare, organizzare e interpretare dati esistenti. Prendere decisioni basate sui dati.
4. Gestione: Pianificare e gestire i dati, includendo l'organizzazione e l'analisi, i protocolli di sicurezza per il salvataggio di dati, la condivisione e la documentazione basata su dati.
5. Comunicazione: Sintetizzare, creare visualizzazioni e rappresentazione di dati.
6. Uso Etico: Identificare fonti di dati diversificati, in particolare dati provenienti dall'attività umana e sociale, considerando i rischi della gestione di tali dati. Comprendere le problematiche implicite all'uso di dati.
7. Preservazione: Essere consapevoli della cura e le pratiche a lungo termine di immagazzinamento, uso e riuso di dati.

Gli approcci non sono stati formulati dal mondo della ricerca; dal movimento sull'Open Data, sono emerse letture sulle competenze necessarie per operare nel settore. La fig.2 presenta il framework "Open Data Skills", generato dal Open Data Institute (ODI)<sup>22</sup>, organizzazione promotrice di formazione avanzata per le aree della data science e le skills professionali di base.

Il framework era stato lanciato focalizzando i livelli del lavoro con dati:

**Esploratore**, chi ha una comprensione basilare dei dati. L'operatore a questo livello è in grado di definire i dati con cui lavorare, indicare esempi o studi di casi e spiegare come possono essere utilizzati i dati per creare il cambiamento.

**Professionista**, chi esegue operazioni di base su un set di dati aperto. È in grado di navigare sui dati e conosce gli strumenti e le tecniche necessari per gestire e pubblicare un set di dati aperto.

**Stratega**, chi integra dati aperti in una strategia o gestisce un progetto di dati in aperto. Conosce le tecniche di pianificazione e di gestione per portare avanti un'iniziativa open data e comprende le sfide inerenti a questo processo.

**Pioniere**, chi possiede abilità e conoscenze che consentono di risolvere sfide del proprio settore. Possono indicare studi di casi specifici per settore, identificare le tendenze future del settore e capire le modalità di uso dei dati che meglio si adeguino alle sfide produttive specifiche del proprio settore.

In una seconda fase, l'ODI ha prodotto un framework interattivo dove va oltre i livelli di competenza per focalizzare le aree e contenuti delle professionalità *data scientist*. Individua come tali l'introduzione all'uso dei dati, pubblicazione, gestione, sviluppo produttivo o *business*, analisi, *leadership*. Tuttavia per l'ODI tali competenze potrebbero essere patrimonio di un'organizzazione, non necessariamente di un lavoratore unico, quindi nella propria offerta formativa e il relativo framework si inquadra un piano di sviluppo organizzativo come "*data literate organization*" piuttosto che come indicazione per il singolo.

sponde ad un determinato membro del dataset in questione (...) Il termine dataset può essere usato anche più genericamente, per indicare i dati in un insieme di tabelle strettamente connesse, relative ad un particolare esperimento o evento (cfr. Wikipedia, definizione consultata in Settembre 2017 <https://it.wikipedia.org/wiki/Dataset>).

22 <https://theodi.org/open-data-skills-framework>, Settembre 2017



Figura 2 – “Open Data Skills Framework” (Open Data Institute)

Come si può inferire questo framework fa riferimento a competenze professionali, che possono creare uno scenario prospettico per lo sviluppo della *data literacy* sin dai primissimi passi dell’istruzione, prendendo appieno come target, invece, le iniziative di formazione continua. Appare però necessario proporre delle basi di lavoro per l’istruzione formale, e quindi per la formazione degli insegnanti e formatori.

In questo esercizio, potremmo introdurre il dibattito europeo sul framework per la Competenza Digitale, DigComp 2.1. (Carretero, Vuorikari, & Punie, 2017). Il concetto, già presente nel primo inquadramento delle cosiddette “competenze chiave di cittadinanza” del 2006 (European Commission, 2007), invita ad una maggiore focalizzazione sul elemento offerto dai contesti e strumenti dell’era digitale, e quindi a pensare ad una serie di conoscenze e abilità pervasive per tutte le altre operazioni della vita quotidiana e professionale. Il primo framework DIGCOMP era nato nel contesto delle operazioni dell’Agenzia Digitale Europea 2020 e rispondeva alle ingenti necessità di qualificare e di generare ambienti qualificanti sul digitale per tutti i cittadini europei di tutte le età. Una nuova versione viene rilasciata nel 2016 (2.0); tuttavia il cambiamento che ci interessa maggiormente è quello dell’ultimo aggiornamento del 2017 (2.1) nel quale alla voce “Information Literacy”, si affianca la voce “Data Literacy”. Il nuovo framework offre non solo la definizione delle competenze, ma apporta una serie di scenari entro cui tali competenze debbono essere applicate e che ne dimostrano il possesso. Siffatta struttura dovrebbe consentire quindi la valutazione della competenza digitale, nelle sue 8 componenti, tra cui l’alfabetizzazione informativa e ai dati. L’attenta lettura dell’intero modello e di questa sottocomponente, nonché delle situazioni considerate per la sua valutazione, sembrano però insufficienti per racchiudere tutta la ricchezza della *data literacy*, nelle componenti che abbiamo prima descritto. Infatti, in DigComp 2.1. parliamo di:

1. Navigare, ricercare e filtrare dati, informazioni e contenuti digitali (nella versione DigComp 1.0: Navigare, ricercare e filtrare le informazioni)

2. Valutare dati, informazioni e contenuti digitali (nella versione DigComp 1.0: Valutare le informazioni)
3. Gestire dati, informazioni e contenuti digitali (nella versione DigComp 1.0: Memorizzare e recuperare le informazioni)

In queste tre sottocomponenti sembra mancare interamente l'aspetto etico, creativo e costruttivo della *data literacy* menzionato appunto in tali modelli.

Se poi DigComp 2.1. offre effettivamente successive aree di analisi che potrebbero offrire uno spazio alle forme più creative della *data literacy*, tale componente non è propriamente identificata e si perde tra le molte altre attività che vanno considerate nelle successive dimensioni. Facendo l'esercizio di allineare dimensioni DigComp con elementi della *data literacy* già analizzati, troviamo che la dimensione comunicazione e collaborazione può trovare un esempio applicato alla *data literacy* nei processi documentazione e narrazione con i dati, ossia *data storytelling*; la dimensione creazione di contenuti digitali, nelle visualizzazioni con dati; la dimensione sicurezza nell'anonimizzazione dei dati; e infine la dimensione problem solving nell'estrazioni complesse ed elaborazioni statistiche.

Questi elementi pongono le basi per pensare ad un framework di competenze a supporto dell'alfabetizzazione ai dati che sebbene risulta convergente con la DigComp, deve trovare il suo spazio, rispondente al fabbisogno formativo che abbiamo a lungo discusso nei paragrafi iniziali.

### *Per un framework di competenze*

Avendo considerato gli sviluppi della *data literacy*, potremmo ora puntare a proporre le basi per un dibattito su un necessario framework di abilità e conoscenze aderenti ai nuovi scenari Big e Open Data.

I traguardi formativi di tipo più generico in relazione ai *data literacy* sono:

- Saper collegare problemi del ambito disciplinare nel quale si lavora o si apprende, con il bisogno di evidenza empirica basata in dati esistenti, da estrarre o da generare.
- Saper distinguere la diversa natura dei dati.
- Saper navigare attraverso informazioni basate su dati, interpretando correttamente grafici e tabelle.
- Saper riconoscere situazioni in cui i dati vengono utilizzati in modo inadeguato, e in particolar modo quanto possono essere stati manipolati per supportare opinione o concetti tendenziosi.
- Riconoscere la differenza della comunicazione basata su dati in relazione ai target comunicativi a cui si rivolge.
- Conoscere strumenti di base per trattare i dati.
- Saper comunicare in modo efficace attraverso l'uso di dati, e con relazione al target a cui si rivolge la comunicazione.

Mentre invece, le competenze più specifiche sono:

- Saper distinguere la diversa natura dei dati, riconoscendo le architetture di dataset e database in base ai problemi esplorati.
- Conoscere e saper utilizzare strumenti e approcci di modellizzazione statistica per l'analisi dei dati.

- Conoscere e saper utilizzare strumenti e approcci per la costruzione avanzata e interattiva (e.g., webapps) di forme di visualizzazione dei dati elaborati.
- Conoscere e saper utilizzare le opportunità offerte dai *big* e *open data* come base alla generazione di servizi e prodotti.

Questi traguardi pongono le basi per un framework per l'alfabetizzazione di base a livello scolastico e di formazione continua, che possa essere utilizzato sia a livello di strutturazione di interventi formativi, sia a livello di valutazione d'impatto di progetti e programmi per lo sviluppo della *data literacy*. La tabella 1 introduce un framework che è frutto dell'elaborazione dell'autrice sulla base dei vari modelli analizzati; si trova in fase di validazione ed è stato sottoposto ad un'iniziale discussione con 5 insegnanti della scuola secondaria di primo e secondo livello tra Luglio e Settembre 2017. Dalla discussione è emersa la necessità di ulteriore declinazione operativa per i vari livelli scolastici, ma le aree risultano trasversali e aderenti alle varie attività in ambito matematico e scientifico. Tuttavia, la discussione con 2 ulteriori insegnanti dell'area umanistica ha consentito di individuare la fattibilità applicativa in progetti interdisciplinari che vedono come oggetto principale le tematiche dell'ambito socio-umanistico, e l'identificazione di dati, nonché loro uso, come azioni sussidiarie.

Dimensioni della competenza [DiGComp]	Macro-descrittori	Micro-descrittori					
			Nulla	Iniziale	Intermedio	Avanzato	Specializzato
Informazione	Ricerca critica di dati	Attivare il pensiero critico in base alle informazioni provenienti dalle attività di analisi dei dati; Conoscere strumenti per la ricerca e identificazione di dati a diverso livello (dal dato crudo ai portali open data); Discriminare i dati in base ad un utilizzo mirato e orientato dai propri obiettivi analitici; Interpretare correttamente grafici e tabelle; Riconoscere quando i dati vengono manomessi, travisati e utilizzati in maniera fuorviante;					
Comunicazione collaborazione	Comunicazione dei dati	Generare datasets come base alla comunicazione dei dati.					
	Collaborazione	Utilizzare strumenti digitali per la gestione e manipolazione congiunta di dati Riconoscere e utilizzare gli ambienti per la condivisione di open dataset					
Creazione di contenuto digitale	Rappresentazione <i>Data storytelling</i>	Creare interfacce per l'interrogazione di grandi <i>dataset</i> Generare visualizzazioni significative con uso di organizzatori grafici, strumenti digitali e statistici. Narrare storie (di significatività culturale/personale) con l'uso di dati. Generare innovazione (servizi/prodotti) attraverso l'uso dei dati.					
Sicurezza	Gestire i dati	Utilizzare correttamente ed eticamente datasets esistenti. Salvare e proteggere i dati.					
Problem solving	Estrarre dati	Usare le tecnologie appropriate per lo scaricamento, salvataggio e lettura di dati. Conoscere strategie e strumenti per il <i>data mining</i> .					

Tabella 1 – Dimensioni e Micro-descrittori della Data Literacy

I livelli di expertise introdotti, formulati in prima persona dovrebbero consentire l'auto ed eterovalutazione. Inoltre, vi è uno spazio dove il valutatore (sia la persona interessata stessa, sia il docente) può indicare esempi di come la competenza e livello di competenza individuata è stata messa in atto. La tabella 2 introduce lo strumento.

Dimensioni	Nulla	Base	Intermedio	Avanzato	Specializzato
<b>Ricerca criticamente i dati</b> [scegli una sola opzione]	Non ho conoscenza o competenza sufficiente per indicare alcuna opzione.	Riconosco l'importanza dei dati per la ricerca e la comunicazione. Riconosco le aree di un mio progetto di lavoro che hanno bisogno di dati.	So individuare i tipi di dati, nonché loro visualizzazioni (grafici, tabelle, schemi) che posso usare per illustrare il mio progetto.	So leggere grafici e tabelle prodotti da altri studenti. Riesco a valutarne l'adeguatezza per quanto riguarda un progetto	So analizzare gruppi di grafici e tabelle. Riesco a valutarne criticamente l'adeguatezza per quanto riguarda ad una comunicazione scientifica, storica, sociale, ecc
<i>Indica un esempio</i>	XXX				
<b>Estrarre i dati</b> [scegli una sola opzione]	Non ho conoscenza o competenza sufficiente per indicare alcuna opzione.	So che i dati possono essere trovati in rete e scaricati per uso personalizzati. So navigare e individuare i tipi di contenuto digitale (e.g. File) dove si trovano i dati che mi servono	So accedere al contenuto digitale dove si trovano i dati che mi servono. So preparare i dati come dataset per l'analisi ulteriore.	So trovare diverso tipo di contenuto digitale. So integrare il contenuto rielaborato in dataset per predisporlo a varie forme di analisi ulteriore.	So elaborare query per la ricerca specifica di contenuto digitale. So scegliere e integrare il contenuto trovato in uno o più dataset per l'analisi ulteriore.
<i>Indica un Esempio</i>	XXX				
<b>Gestire i dati</b> [scegli una sola opzione]	Non ho conoscenza o competenza sufficiente per indicare alcuna opzione.	So che è importante organizzare i dati e proteggerli. So usare strumenti digitali di base in locale per organizzare dati trovati in rete.	So usare strumenti digitali di base in locale per organizzare e proteggere dati trovati in rete. So individuare dataset che contengono dati sensibili.	So usare strumenti digitali di base in locale oppure sul cloud per organizzare e proteggere dati trovati in rete. So anonimizzare dati sensibili in un dataset.	So usare strumenti digitali di base in locale oppure sul cloud per organizzare e proteggere dati trovati in rete. So rielaborare dati sensibili in un dataset per renderli aperti.
<i>Indica un esempio</i>	XXX				
<b>Comunicare e collaborare</b> [scegli una sola opzione]	Non ho conoscenza o competenza sufficiente per indicare alcuna opzione.	So che esistono ambienti digitali dove i dati possono essere analizzati in modo collaborativo. So indicare ad altri compagni di gruppo dove possono trovare dati utili ad un progetto in corso.	So individuare ambienti web, social o digital library dove si trovano dati aperti. Riesco ad individuare strumenti di comunicazione offerti per condividere informazioni o fare domande sui dati.	So partecipare (isciversi, lasciare un post su forum, ecc.) su ambienti web, social o digital library dove si trovano dati aperti. Riesco a condividere i miei dati con altri compagni che conosco.	Sono stato/sono attivo su qualche ambiente web, social o digital library dove si trovano dati aperti. Riesco a condividere i miei dati aperti con altri membri della comunità digitale. Riesco ad aprire i miei dati per la collaborazione ulteriore.

<b>Indica un esempio</b>	XXX				
<b>Raccontare con i dati [scegli una sola opzione]</b>	Non ho conoscenza o competenza sufficiente per indicare alcuna opzione.	So che è importante usare criticamente i dati per fondamentare la propria comunicazione. Riesco a usare strumenti digitali (app e software) per creare visualizzazioni su miei dati. So usare il linguaggio statistico e gli organizzatori grafici di base per le visualizzazioni.	Riesco a integrare visualizzazioni create con strumenti digitali (app e software) in rappresentazioni infografiche. So usare software e app per le rappresentazioni infografiche digitali. So usare il linguaggio statistico e gli organizzatori grafici avanzati per le visualizzazioni.	So usare software e app per la visualizzazione e la creazione di rappresentazioni infografiche con elementi statistici, numerici o di schematizzazione grafica interattiva.	So creare strumenti/interfacce digitali avanzate per la visualizzazione di rappresentazioni infografiche con elementi statistici, numerici o di schematizzazione grafica interattiva.
<b>Indica un esempio</b>	XXX				

**Tabella 2 – Strumento per la etero e co-valutazione della Data Literacy**

Lo strumento, disponibile online, può essere usato, condiviso, riadattato (licenza CC-BY 4.0) ed in fase di sviluppo verso forme più sofisticate ed interattive (feed-back per l'attività formativa al insegnante/formatore in seguito all'uso diagnostico –quanto sanno gli studenti in entrata)<sup>23</sup>.

### Riflessioni conclusive

In questo contributo l'obiettivo è stato quello di aprire un dibattito pedagogico relativamente ad un fenomeno emergente che sta mobilitando interesse e risorse, generando innumerevoli opportunità di sviluppo sia economico, sia culturale. La tesi centrale dalla quale è partito il presente lavoro è stata che per appropriarsi del tesoro che racchiudono i *big e open data*, risulta necessario proporre una base analitica che orienti la ricerca educativa e la formazione dei formatori in materia di *data literacy*, competenza necessaria per operare nei suddetti nuovi contesti.

Dopo una sintetica e sicuramente incompleta disamina del fenomeno *big-open-data*, volta a supportare la comprensione dello stesso da parte della comunità di ricerca e pratica educativa alla quale questo lavoro si rivolge, il focus si è spostato sulla definizione di *data literacy*, discutendo ulteriormente alcuni framework per la sua analisi e sviluppo. In particolare, sono stati discussi due punti: la necessità di andare oltre il framework "DigComp" per la copertura specifica della *data literacy* (nonostante l'importantissimo punto di partenza che esso fornisce); e l'inadeguatezza di alcuni framework esistenti per caratterizzare l'al-

<sup>23</sup> Questionario per l' Auto-valutazione: <https://goo.gl/forms/3uQWz5lJfARndz1S2>

Questionario per l'etero e co-valutazione: <https://goo.gl/forms/2LdtdOQfptJD3ruC3>

fabetizzazione ai dati dalla base del sistema dell'istruzione e della formazione, non solo come competenza specialistica.

Infine, è stato presentato un lavoro in via del tutto prototipale, che muovendo dall'analisi concettuale e un *primus inter pares* sulla fattibilità applicativa, propone uno strumento utile alla pratica e la ricerca educativa sulla *data literacy*. Di validità ancora da consolidare vista la modalità di elaborazione, tenta di porre le basi per un progressivo perfezionamento che l'autrice e chi ne sia interessato nella comunità di ricerca e pratica educativa potrebbe svolgere a supporto di un tema che, come nel contesto internazionale, richiede sicura attenzione.

Molto resta ancora da fare, dal punto di vista della ricerca didattica e la ricerca educativa di base. Oltre la validazione di un framework a livello nazionale, risultano necessari numerosi interventi per orientare ulteriormente l'impegno che il MIUR propone già come orientamento.

In primo luogo, nonostante la raccolta di evidenza del rapporto MIUR del gruppo Big Data, e quella introdotta in questa sede, è evidente che sia necessario ancora mappare il fabbisogno formativo e le pratiche esistenti, a partire da approcci di ricerca esplorativa di tipo ampio e randomico (indagini nazionali e regionali a più livelli di istruzione); ad approcci di tipo qualitativo in cui il vissuto di esperti *data scientist*, tra altri stakeholder, vengono esplorati in profondità.

In secondo luogo, partendo dalla base orientativa fornita dal MIUR, sarebbe possibile e desiderabile implementare forme di ricerca su progetto (*design-based research*), a partire da iniziative innovative che accompagnano la formazione degli insegnanti e docenti universitari. Su questa base, proporre la *ricerca didattica*, come analisi riflessiva e documentata condotta da insegnanti e docenti universitari sulle proprie classi, a partire dalla sperimentazione di approcci diversificati di alfabetizzazione ai dati, oppure all'introduzione di moduli sussidiari alla formazione di altre competenze. In particolare, sarebbe importante focalizzare l'innovazione didattica sulle seguenti linee:

- Per la scuola primaria, è necessario focalizzarsi su skill a livello molecolare, come per esempio la comprensione e uso di organizzatori grafici, la nozione di "evidenza", giocare a mettersi nei panni di uno scienziato per capire il processo di raccolta informazioni e di elaborazione basate su operazioni matematiche proprie del livello in questione. Sono fattibili inoltre le attività di sviluppo di infografiche, prima cartacee e poi digitali.
- Per la scuola secondaria di I e II livello, sono proposte interessanti e trasversali la narrazione di storie basate su dati, le *data expeditions* per esplorare la scienza aperta e partecipata, e la cittadinanza attiva attraverso l'uso di dati aperti.
- A livello universitario, si può trasversalmente dare luogo, in assolutamente tutte le discipline, all'uso di dati aperti come risorse didattiche aperte. In questo modo, il pensiero scientifico si apre non solo nella sua fase conclusiva ma anche nelle fasi processuali, ponendo le basi per una *data literacy* che sarà poi spendibile sia a livello professionale, sia a livello accademico. In particolare, sarebbe desiderabile attivare progetti interdisciplinari e interdipartimentali nei quali dipartimenti di statistica e informatica collaborano con dipartimenti di aree sociali e umanistiche per lo sviluppo di servizi/applicativi a partire dall'uso di Open Data, come compito autentico e situato.

Ulteriormente, la ricerca educativa dovrebbe puntare al collaudo di framework per la *data literacy* (con specificità a livello d'istruzione e formazione). Non meno importante, sarebbe utile sviluppare ambienti e tools digitali (simulazioni,

games, hackaton, applicativi) a supporto di ambienti e azioni formative alla *data literacy*. Inoltre, sarebbe fondamentale considerare approcci critici, di riflessione epistemologiche che vedono l'intelligenza artificiale (cui principi e strumenti si usano nei dispositivi automatizzati di elaborazione dati (*analytics, text mining, natural language processing, machine learning*) e le sfide etiche che tali dispositivi pongono alla società. La comunità scientifica più estesa deve riconoscere la natura assolutamente interdisciplinare degli approcci di studio sui *big e open data*. La caratterizzazione del problema mette in atto sicuramente in primo luogo una tecnicità informatica e statistica, nonché delle scienze dell'informazione, ma non può prescindere dalla caratterizzazione dei problemi sociali e culturali che solo le scienze umanistico-sociali possono fare; e men che meno, di un dibattito filosofico e critico sull'ontologia e le epistemologie sottostanti alla caratterizzazione dei dati, concetto assai cospicuo, come abbiamo afferrato all'inizio di questo lavoro.

Come si può evincere da queste prospettive, il gap tra scenario di innovazione e scenario formativo è di difficile definizione. Il problema formativo a più livelli, resta quello di integrare prospettive emergenti di formazione, consolidare strumenti concettuali e operativi, sperimentare processi, e misurarne gli impatti. Si consideri in questa chiusura, quasi quale provocazione all'approfondimento della prospettiva l'impianto teorico fornito da un gruppo di pedagogisti studiosi delle ricadute del pensiero di Amartya Sen nella riformulazione di Martha Nussbaum (Alessandrini, 2014; Costa, 2013; Umberto Margiotta, 2011). Per questo gruppo i contesti di azione socio-culturale possono essere modulati attraverso il progetto formativo per diventare "capacitanti": contesti che offrono possibilità di sviluppo delle capacità (*capacity*) intese nel doppio senso di ciò che la persona è in grado di fare ma anche ciò che la persona può potenzialmente fare in un contesto che apra a diverse opportunità (*functionings*). Così, la persona costruisce un senso di agentività, di progetto di sé in funzione di un processo di sviluppo socio-culturale. In linea con questi presupposti, il fenomeno dei *Big e Open Data* possono costituire un contesto (digitale) capacitante; i vari dispositivi formativi, nel senso dei modelli, strumenti metodologie educative per una *alfabetizzazione ai dati*, il modo con cui il fenomeno diventa trama enattiva di nuove competenze, appropriazione e costruzione di senso verso un processo ricorsivo di democratizzazione della conoscenza.

Ma la storia, ad oggi, è tutta da scrivere.

## References

- Alessandrini, G. (2014). *La pedagogia di Martha Nussbaum: approccio alle capacità e sfide educative*. F. Angeli.
- Atenas, J., Havemann, L., & Priego, E. (2015). Open Data as Open Educational Resources: Towards Transversal Skills and Global Citizenship. *Open Praxis*, 7(4), 377–389. <http://doi.org/10.5944/openpraxis.7.4.233>
- Ayris, P., Berthou, J.-Y., Bruce, R., Lindstaedt, S., Monreale, A., Mons, B., ... Tochtermann, Klaus Wilkinson, R. (2016). *Realising the European Open Science Cloud. First report and recommendations of the Commission High Level Expert Group on the European Open Science Cloud*. Luxembourg.
- Baack, S. (2015). Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism. *Big Data & Society*, 2(2), 205395171559463. <http://doi.org/10.1177/2053951715594634>
- Carlson, J., Fosmire, M., Miller, C. C., & Nelson, M. S. (2011). Determining Data Information Literacy Needs: A Study of Students and Research Faculty. *Portal: Libraries and the*

- Academy, 11(2), 629–657. <http://doi.org/10.1353/pla.2011.0022>
- Carretero, S., Vuorikari, R., & Punie, Y. (2017). *The Digital Competence Framework for Citizens With eight proficiency levels and examples of use*. Brussels. <http://doi.org/10.2760/38842>
- Costa, M. (2013). Forma-Azione: i processi di capacitazione nei contesti di innovazione Educ(a)ction: Capability processes within contexts of innovation. *Formazione&Insegnamento.*, XI(1), 103–118. Retrieved from [http://www.formazione-insegnamento.net/files/FORMAZIONE&INSEGNAMENTO\\_1-2013/FORMAZIONE&INSEGNAMENTO\\_1-2013.pdf](http://www.formazione-insegnamento.net/files/FORMAZIONE&INSEGNAMENTO_1-2013/FORMAZIONE&INSEGNAMENTO_1-2013.pdf)
- Davies, T. (2010). Open data , democracy and public sector. *Interface*, 1–47. Retrieved from <http://practicalparticipation.co.uk/odi/report/wp-content/uploads/2010/08/How-is-open-government-data-being-used-in-practice.pdf>
- EMC Education Services. (2015). *Data Science & Big Data Analytics*. Indianapolis, IN, USA: John Wiley & Sons, Inc. <http://doi.org/10.1002/9781119183686>
- European Commission. (2007). *Key Competences for Lifelong Learning. European Reference Framework*. Luxembourg. Retrieved from [http://ec.europa.eu/dgs/education\\_culture/publ/pdf/ll-learning/keycomp\\_en.pdf](http://ec.europa.eu/dgs/education_culture/publ/pdf/ll-learning/keycomp_en.pdf)
- European Commission. (2013). *Digital science in Horizon 2020. European Commission*. Brussels. Retrieved from [ec.europa.eu/information\\_society/newsroom/cf/dae/document.cfm?doc\\_id=2124](http://ec.europa.eu/information_society/newsroom/cf/dae/document.cfm?doc_id=2124)
- European Commission. (2015). *Consultation on “Science 2.0”: Science in Transition*. Brussels. Retrieved from [https://ec.europa.eu/research/consultations/science-2.0/consultation\\_en.htm](https://ec.europa.eu/research/consultations/science-2.0/consultation_en.htm)
- European Commission. (2016a). *Open innovation, open science, open to the world - a vision for Europe | Digital Single Market*. Luxembourg. Retrieved from <https://ec.europa.eu/digital-single-market/en/news/open-innovation-open-science-open-world-vision-europe>
- European Commission. (2016b). Open Science Monitor. Retrieved September 25, 2017, from <http://ec.europa.eu/research/openscience/index.cfm?pg=home&section=monitor>
- European Commission Decision. (2016). Horizon 2020 Work Programme 2016 - 2017. 16. Science with and for Society, 2017(July 2016), 1–72. Retrieved from [http://ec.europa.eu/research/participants/data/ref/h2020/wp/2016\\_2017/main/h2020-wp1617-swfs\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/wp/2016_2017/main/h2020-wp1617-swfs_en.pdf)
- Gartner Inc. (2013). What Is Big Data? - Gartner IT Glossary - Big Data. Retrieved from <https://research.gartner.com/definition-what-is-big-data?resId=3002918&srcId=1-8163325102>
- Gruppo di lavoro MIUR sui Big Data. (2016). *Big Data @MIUR. Rapporto del gruppo di lavoro*. Roma. Retrieved from <http://www.istruzione.it/allegati/2016/bigdata.pdf>
- Gurstein, M. B. (2011). (2011). Open data: Empowering the empowered or effective data use for everyone? | Gurstein | First Monday. *First Monday*, 16(2), 1–8. <http://doi.org/10.1177/0170840601223003>
- Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*, 29(4), 258–268. <http://doi.org/10.1080/10580530.2012.716740>
- Johnson, P., & Robinson, P. (2014). Civic Hackathons: Innovation, Procurement, or Civic Engagement? *Review of Policy Research*, 31(4), 349–357. <http://doi.org/10.1111/ropr.12074>
- Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1). <http://doi.org/10.1177/2053951714528481>
- Konstantopoulos, S., & Traynor, A. (2014). Class size effects on reading achievement using PIRLS data: Evidence from greece. *Teachers College Record*, 116(2).
- Lathrop, D., & Ruma, L. (2010). *Open Government Collaboration, Transparency, and Participation in Practice*. O’Reilly Media.
- Lohr, S. (2013). The Origins of “Big Data”: An Etymological Detective Story. *New York Times*, online. Retrieved from <https://bits.blogs.nytimes.com/2013/02/01/the-origins-of-big-data-an-etymological-detective-story/>
- Lupton, M., & Bruce, C. S. (2010). Windows on information literacy worlds: Generic, situated and transformative perspectives. In A. Lloyd & S. Talja (Eds.), *Practicing information literacy: Bringing theories of learning, practice and information literacy together* (pp.

- 4–27). Wagga-Wagga: Centre for Information Studies, Charles Sturt University.
- Manca, S., & Raffaghelli, J. E. (2017). Towards a multilevel framework for analysing academic social network sites: A networked socio-Technical perspective. In *Proceedings of the 4th European Conference on Social Media, ECSM 2017*.
- Matt, S. (2016). e-Infrastructures to facilitate Open Scholarship. Retrieved January 5, 2017, from <https://ec.europa.eu/futurium/en/content/e-infrastructures-facilitate-open-scholarship>
- Maybee, C., & Zilinski, L. (2015). Data informed learning: A next phase data literacy framework for higher education. *Proceedings of the Association for Information Science and Technology*, 52(1), 1–4. <http://doi.org/10.1002/pra2.2015.1450520100108>
- OKF - Open Knowledge Foundation. (n.d.). Definizione di Conoscenza Aperta. Retrieved September 25, 2017, from <http://opendefinition.org/od/2.0/it/>
- Open Knowledge International. (2014). The Open Data Handbook. Retrieved September 25, 2017, from <http://opendatahandbook.org/>
- Raffaghelli, J. E. (2014). A Scholarship of Open Teaching and Learning: new basis for quality in higher education. *Formazione&Insegnamento, European Journal of Research on Education and Teaching*, 12(1), 211–244. Retrieved from <http://ojs.pensamultimedia.it/index.php/siref/article/view/380>
- Raffaghelli, J. E. (2017). Exploring the (missed) connections between digital scholarship and faculty development: a conceptual analysis. *International Journal of Educational Technology in Higher Education*, 14(1), 20. <http://doi.org/10.1186/s41239-017-0058-x>
- Ridsdale, C., Rothwell, J., Smit, M., Ali-Hassan, H., Bliemel, M., Irvine, D., ... Wuetherick, B. (2015). *Strategies and Best Practices for Data Literacy Education*. Retrieved from <http://dataliteracy.ca/>
- Smolan, R., & Erwitte, J. (2012). The Human Face of Big Data. *Science*, 351(6274), 223. <http://doi.org/10.1126/science.aaf3194>
- Stephenson, E., & Schifter Caravello, P. (2007). Incorporating data literacy into undergraduate information literacy programs in the social sciences. *Reference Services Review*, 35(4), 525–540. <http://doi.org/10.1108/00907320710838354>
- Umberto Margiotta. (2011). Capitale formativo e welfare delle persone. Verso un nuovo contratto sociale. *MeTis. Mondì Educativi, Temi, Indagini, Suggestioni.*, 37(2), online. Retrieved from <http://www.metis.progedit.com/home/37-saggi/191-capitale-formativo-e-welfare-delle-persone-verso-un-nuovo-contratto-sociale.html>
- Zuiderwijk, A., & Janssen, M. (2014). Open data policies, their implementation and impact: A framework for comparison. *Government Information Quarterly*, 31(1), 17–29. <http://doi.org/10.1016/j.giq.2013.04.003>
- Zuiderwijk, A., Janssen, M., Choenni, S., Meijer, R., & Alibaks, R. S. (2012). Socio-technical Impediments of Open Data. *Electronic Journal of E-Government*, 10(2), 156–172. <http://doi.org/10.1641/b570402?ref=search-gateway:885882d1830675b0f27af0760faeaf8>