

Come il linguaggio dei big data cambia la comunicazione e l'accesso alla conoscenza

Elena Guerra

Università degli studi di Brescia -
Dottorato di ricerca in ingegneria meccanica e industriale

1. Introduzione

Negli ultimi decenni, la progressione del sapere tecnico-scientifico ha generato una simultanea mole di dati digitali eterogenei in crescita esponenziale, i cosiddetti big data, che ha modificato e sta continuamente modificando l'approccio alla conoscenza, la sua stessa configurazione, nonché la trasmissione del sapere. Generalmente, senza la mediazione delle macchine virtuali (VM) – software che hanno le stesse funzionalità dei computer fisici – oggi non è più possibile, non solo produrre dati che siano portatori di informazioni, ma neppure organizzare digitalmente quegli stessi dati in maniera sistematica per consentire il loro successivo ritrovamento e una nuova fruizione. La stessa operazione di recupero dei dati – archiviati nei *database*, i contenitori virtuali di dati digitali – presuppone l'ausilio di macchine che compiono l'operazione di ricerca ed estrapolazione per poi consentirne, magari attraverso ulteriori macchine, la lettura e l'interpretazione (Budroni, 2021a; Budroni, 2021b). I dati digitali, infatti, portano con sé, oltre alle informazioni di contenuto, altre informazioni aggiuntive, i cosiddetti *metadati*, che ne consentono l'identificazione attraverso una accurata descrizione, utile al loro ritrovamento nel *mare magnum* dei dati digitali.

L'uso o il riuso dei dati, inoltre, può essere sottoposto o meno a dei vincoli, che ne consentano l'impiego nei diversi contesti. In questo caso si parla di *open data*, *non open data*, *restricted access data*. A seconda del contesto di utilizzo e per motivi legali, i dati, infatti, possono essere ad accesso aperto o chiuso, oppure prevedere un accesso ristretto. La produzione di dati, l'archiviazione e le nuove ricerche per successive condivisioni e riutilizzi sono dunque operazioni sempre mediate dall'uso delle macchine digitali, che parlano tra loro attraverso un linguaggio informatico ricco di anglicismi.

Normalmente, la padronanza di questo linguaggio è prerogativa di un gruppo ristretto di tecnici specializzati, i cosiddetti *data scientist*, gli scienziati dei dati.

Il termine fu coniato da D.J. Patil e Jeff Hemmerbacher, i primi pionieristici iniziatori di questa disciplina, che, a partire dal 2008, cominciarono ad utilizzarlo per indicare quegli analisti che, non solo predispongono i dati per essere analizzati ed esplorati, ma costruiscono anche modelli di dati utilizzando diversi linguaggi di programmazione e li configurano per governarli. Oggi, migliorare la configurazione di questi modelli consente anche una gestione più efficace ed efficiente della conoscenza e della sua condivisione (Xiao *et al.*, 2020; Budroni *et al.*, 2023).

Sul piano linguistico e culturale stiamo assistendo a un progressivo e rivoluzionario fenomeno di contaminazione: il linguaggio tecnico-informatico è uscito dall'*hortus conclusus* in cui operano gli addetti ai lavori e sta contaminando il linguaggio naturale dei parlanti le diverse lingue del mondo. Accedere al linguaggio e utilizzare correttamente alcuni termini specifici consente, infatti, di collegarli ad azioni altrettanto specifiche. Questa procedura è diventata la prima chiave d'accesso all'espletamento di azioni ordinarie nel quotidiano, tanto che, per compierle correttamente e comprendere il linguaggio tecnico-informatico, sono stati approntati anche dei dizionari enciclopedici (Cancila, Mazzanti, 2009). Ciascuno di noi, costantemente, paga la spesa con un bancomat, il pedaggio autostradale con il telepass, ricerca una meta turistica in internet o invia un messaggio con lo smartphone. Ciò implica un utilizzo di dati digitali, a prescindere dalla consapevolezza che abbiamo dei meccanismi che sono ad essi sottesi. La vita stessa, in gran parte della società globalizzata, è così correlata al digitale e alle operazioni ad esso connesse che non c'è più soluzione di continuità tra reale e virtuale, al punto che: «tutto ciò che è reale è digitale e viceversa» (Guerra, 2021). La complessità del mondo contemporaneo presuppone, quindi, oltre allo sviluppo di competenze specialistiche molteplici, una sempre più esplicita padronanza di linguaggi specialistici come chiave d'accesso a queste competenze che si intersecano fra loro e i cui confini disciplinari confluiscono gli uni negli altri in maniera sempre più sfumata. L'approccio ai problemi, infatti, non è solo teorico, ma spesso si coniuga con capacità pratiche e anche linguistiche e comunicative. Se, banalmente, voglio produrre un documento e inviarlo in formato digitale per condividerlo con un gruppo ristretto di utilizzatori, dovrò possedere, non solo le competenze disciplinari per stilarlo, ma anche le competenze informatiche per inviarlo correttamente, quelle linguistiche utili a comprendere l'inglese del gergo tecnico-informatico (che non è quello che i parlanti utilizzano correntemente) e quelle legali per apporre le opportune restrizioni di utilizzo. I flussi di dati che

Come il linguaggio dei big data cambia la comunicazione e l'accesso alla conoscenza

irrompono nelle nostre vite, dunque, le plasmano e le modificano irreversibilmente e velocemente anche nel modo di comunicare.

Questo articolo si propone di mettere in evidenza alcuni passaggi cruciali di questo fenomeno, ormai irreversibile, tenendo conto di alcuni degli studi più recenti (è stato utilizzato il database Scopus impostato secondo criteri legati alla multidisciplinarietà e a diverse parole chiave, come ad esempio big data e conoscenza) e del confronto culturale che gli obiettivi della Risoluzione del 20 ottobre 2020 del parlamento europeo sugli aspetti etici relativi all'intelligenza artificiale, alla robotica e alle tecnologie ad esse collegate hanno avviato (European Parliament resolution, 20 ottobre 2020).

2. I big data e la trasformazione linguistica e culturale che accompagna la relazione umano-macchina

Durante il periodo della pandemia, tra il 2019 e il 2022, abbiamo vissuto come vitale l'interconnessione digitale, che ha garantito, non solo i contatti e le relazioni fra umani in tutte le comunità del mondo, ma soprattutto la trasmissione, la gestione e la condivisione di quei dati essenziali per la sopravvivenza umana e il controllo dei contagi. In tutto il globo le persone si sono trovate chiuse nelle proprie case o nei luoghi di lavoro e, in maniera pressoché simultanea, dei comportamenti individuali sono diventati comportamenti collettivi. Per essere connessi e poter comunicare fra loro, anche oralmente, gli umani hanno dovuto condividere dei protocolli standardizzati; lo scambio di messaggi, l'ascolto della voce, la visione degli interlocutori sono stati resi possibili dall'uso di macchine come computer o telefoni cellulari. Questa prassi comunicativa, anche una volta finita l'emergenza pandemica, è rimasta in uso, ad esempio, per creare connessioni fra gruppi di lavoro ubicati in paesi diversi, per lo smart working, per rendere più efficienti le comunicazioni a distanza in genere.

In concreto, le informazioni, per poter essere scambiate digitalmente tra gli utenti, vengono trasformate da algoritmi in dati digitali (*data*), corredati da ulteriori dati (*metadata*) che descrivono i dati a cui sono associati e ne consentono, fra l'altro, la codifica e decodifica. La comunicazione è pertanto mediata dall'uso delle macchine (*Human-Machine-Human*) e le informazioni sono possibili grazie al transito di enormi volumi di dati, i big data, attraverso delle autostrade virtuali, denominate *e-infrastructures*, infrastrutture digitali. Tutti questi termini appartengono a un codice linguistico specifico che, nel contesto pandemico, è apparso vitale, anche per i cittadini comuni, conoscere almeno nei suoi elementi basilari.

Molti termini, associati ai dati digitali e ai processi che consentono di estrapolare da questi ulteriori dati di conoscenza, sono, tuttavia, espressi in lingua inglese e appartengono a un gergo tecnico-informatico standardizzato. L'incrocio tra un inglese *pidgin* – il prodotto di una lingua semplificata, generata da parlanti che utilizzano l'inglese per risolvere problemi di comunicazione, così da rendere il più agevole possibile lo scambio di messaggi – e un gergo specialistico, fa sì che apprendere questo linguaggio non sia così automatico, proprio per le conoscenze che implica (Guerra, 2021). Senza una prima alfabetizzazione e il possesso di un vocabolario minimo di quella specifica lingua e della sua prassi di condivisione, ogni comunicazione risulta preclusa. In un contesto emergenziale, è evidente, la possibilità di comunicare è spesso salvifica. Molti termini, come *data visiting*, *data sharing*, *data stewardship*, conati negli ultimi 13 anni, sono usciti, prima gradualmente e poi, anche grazie alla pandemia, sempre più rapidamente, dai contesti specialistici e stanno progressivamente diventando di uso comune, proprio perché apprenderli è utile e funzionale nella vita di ogni giorno (Budroni, 2021a). Con l'espressione *data sharing* si intende la condivisione dei dati digitali, ovvero quella tappa di processo attraverso cui passa la condivisione delle conoscenze. Gli studiosi di scienza dell'informazione Chawinga e Zinn, a questo proposito, sottolineano un concetto rilevante che traduciamo dall'inglese:

La condivisione dei dati (*data sharing*) è inevitabile in un mondo moderno dove parte della nuova conoscenza è scoperta attraverso la rianalisi dei dati generati in precedenza. (Chawinga, Zinn, 2019)

Quindi, condividere i dati è un passaggio essenziale per produrre nuova conoscenza, ma parte della nuova conoscenza è prodotta non da nuovi dati, come sarebbe intuitivo pensare, ma da una rielaborazione di dati che già erano stati generati e che emergono per la prima volta o che vengono visti in modo differente rispetto al passato per dirci cose nuove. Prima di poter condividere delle informazioni o delle conoscenze, però, sarà necessario poter accedere ai dati in maniera sicura e controllata, compiendo un'operazione di *data visiting*. In pratica i dati devono poter essere visibili ed anche leggibili. Spesso, le informazioni sono infatti seppellite sotto cumuli enormi di dati e bisogna reperirle. Cosa non facile. Tutto ciò non è un processo immediato, ma anzi, passa attraverso la mediazione delle macchine, che parlano tra loro interfacciandosi. Software, applicazioni, algoritmi, sono vocaboli a cui attinge il loro linguaggio, variegato e composito, che deve essere tradotto per essere più accessibile e comprensibile. Solo con questa decrittazione del linguaggio informatico si potrà accedere alle

Come il linguaggio dei big data cambia la comunicazione e l'accesso alla conoscenza

informazioni di cui sono portatori i dati per poi interpretarli e diffondere la conoscenza ad essi correlata.

In particolare, ciascun utente che volesse ottenere una risposta da un motore di ricerca, per esempio circa la variabilità climatica in una determinata area geografica, potrebbe andare sul sito dell'aeronautica militare, oppure su *copernicus.eu* e qui dovrebbe digitare la sua richiesta, sapendo che, come recita il sito: «la stragrande maggioranza dei dati e delle informazioni provenienti dall'infrastruttura spaziale e dai servizi di Copernicus è disponibile e accessibile gratuitamente per i cittadini e le organizzazioni di tutto il mondo». Il sito, infatti, mette a disposizione i dati attraverso le piattaforme tradizionali di accesso con due punti gestiti, nello specifico, dall'ESA, l'Agenzia Spaziale Europea. In concreto, la ricerca parte tramite dei servizi che vengono messi a disposizione da un'azienda o da un ente. Tramite questi servizi è quindi possibile accedere a dei dati della ricerca che sono stati creati da macchine e che poi, eventualmente, sono stati validati da altri servizi, anch'essi gestiti attraverso ulteriori strumenti digitali, in modo che il responso della ricerca fornirà la variabilità climatica prevista per l'area geografica desiderata. Per ottenere questa risposta, l'utente non è entrato nel *database* dell'ESA e non ha scaricato tutti i dati che riguardano l'area geografica di interesse, ma ha semplicemente *visitato* dei dati. La differenza tra il *data sharing* e il *data visiting*, in questo caso, emerge chiaramente: per visionare non ho bisogno di condividere dati. Questo passaggio mette in luce un altro concetto rilevante: per conoscere non ho necessariamente bisogno di condividere dati, ma è sufficiente poter accedere ai dati e visionarli.

Per poter rinvenire i dati, infine, è necessario che essi siano predisposti digitalmente. A questo punto, entrano in gioco i *data stewards*, termine che identifica quelle figure professionali preposte alla preparazione dei metadati associati ai dati, così da consentirne prima il ritrovamento da parte dei motori di ricerca e poi la fruizione.

3. I motori di ricerca e l'orientamento della conoscenza

È proprio i motori di ricerca ad esercitare un potere sottile che spesso non è percepito dall'utilizzatore finale. In Europa, poiché non possediamo grossi motori di ricerca, il cui sviluppo richiede tecnologie molto avanzate, ci affidiamo a chi offre questi servizi online, per esempio Google, azienda leader nell'innovazione a tutto campo, madre dell'omonimo motore che, nato nel 1998 in California, è oggi il motore più usato al mondo. Gli studiosi Riccardo Finocchi e

Dario Cecchi, a questo proposito, sottolineano che «oltre a mappare l'universo esistente Google tende a sua volta a configurarsi come un universo, locative medium che influisce sulla stessa percezione dello spazio, nell'interazione tra realtà e simulazione, off line e online» (Del Marco, Pezzini, 2017). Ovviamente i grandi competitors commerciali ed economici degli Stati Uniti, come la Cina, hanno sviluppato dei propri motori di ricerca che, a loro volta, prospettano una vera e propria *weltanschauung*, una visione del mondo che irrompe nelle nostre vite. Pensiamo solo alle straordinarie immagini offerte, che conducono lo sguardo oltre lo schermo, modificando le nostre aspettative sulla visione del reale e sull'interazione con esso. Incantati dai vantaggi che i motori di ricerca offrono, dimentichiamo ciò che diamo loro compiendo una ricerca: informazioni che riguardano noi stessi e la nostra vita professionale e privata, ma anche aspirazioni o desideri.

Inviare una richiesta tramite dei motori di ricerca non soltanto produce dei risultati attesi, ma produce ulteriori dati generati dalla ricerca stessa, e a questa ricollegabili. Questi dati risultano molto interessanti per chi li gestisce e possono essere analizzati e utilizzati per diverse finalità.

Profilare l'utente utilizzatore, sulla base, ad esempio, del tipo di richiesta effettuata, dei servizi selezionati, degli interessi culturali o commerciali manifestati, diventa prioritario per tutte le aziende, specie per le grandi compagnie internazionali, interessate ad usare la rete per pubblicizzare o commercializzare prodotti o servizi. In pratica, le tracce lasciate in rete attraverso i click diventano dati, che vengono letti e interpretati dalle aziende, non solo per soddisfare i consumatori ma anche per orientarne i consumi, le idee o le scelte politiche. (Guerra, 2021). Dal punto di vista degli utenti utilizzatori, rimanere padroni dei propri dati significa anche gestire in modo migliore la propria "identità digitale". Quello che gli utenti o le macchine ritrovano in seguito a una ricerca è infatti quello che contribuisce a creare un'identità digitale, che si sovrappone all'identità reale e, nel migliore dei casi, cosa che spesso non accade, coincide con questa (Budroni, 2021b).

4. *Internet delle cose*

Un'ulteriore estensione di Internet al mondo degli oggetti e dei luoghi concreti è costituita dal cosiddetto *Internet of Things* (IoT), che rappresenta un settore prioritario del mercato unico digitale. Gli oggetti comunicano attraverso la rete con altri oggetti – comprese le macchine per l'elaborazione industriale

Come il linguaggio dei big data cambia la comunicazione e l'accesso alla conoscenza

e gli elementi elaborati industrialmente (sistemi cyber-fisici), le attrezzature domestiche, l'elettronica indossabile, i veicoli e i sensori – acquisendo un'identità digitale e sono in grado di fornire servizi agli utenti (Rolling Plan for ICT standardization, 2023). Esempi concreti sono rappresentati dai contatori di gas intelligenti (smart metering) installati presso le utenze domestiche. Secondo il report dell'Osservatorio Internet of Things della School of Management del Politecnico di Milano, presentato nel corso del convegno “Internet of Things: tra continuità e innovazione” del 2022, il mercato degli smart object in Italia è arrivato a toccare gli 8,3 miliardi di fatturato complessivo, registrando un +13% sul 2021. Nel prossimo futuro si prevede un'ulteriore accelerazione del mercato, soprattutto per quanto riguarda gli ambiti *Smart Metering*, *Smart Car*, *Smart Home* e *Industrial IoT*. Secondo il Rolling Plan for ICT standardization 2023 dell'Europa, a livello globale, sarà possibile raggiungere 50 miliardi di dispositivi in uso in tutto il mondo entro il 2030. La massiccia rete di dispositivi interconnessi genererà 79,4 zettabyte (ZB) di dati. Oltre al potenziale di innovazione in molti settori industriali, l'IoT potrà contribuire ad affrontare molte sfide sociali, tra cui i cambiamenti climatici, l'invecchiamento, l'efficienza energetica e delle risorse.

5. *Governance dei dati*

La *governance dei dati*, la gestione dei dati da parte delle compagnie che li detengono, costituisce un ambito di ricerca molto rilevante, che nell'Unione Europea è regolato dal GDPR, il Regolamento generale sulla protezione dei dati, promulgato il 25 maggio 2018. Esso si applica a tutte le imprese con sede nell'UE e a quelle al di fuori dell'UE che offrono beni o servizi a persone che vivono in Paesi UE o comunque che ne monitorano i comportamenti. Poiché tutte le linee di business devono rispettare il GDPR, la *governance dei dati* deve essere gestita in tutti i suoi aspetti, con particolare attenzione alla privacy dei dati (GDPR, 2018). Lo sforzo del legislatore è orientato ad uniformare e semplificare le diversità legislative dei diversi paesi europei in tema di privacy e gestione di dati. Infatti, rendere omogenea la gestione dei dati consente un'interfaccia migliore da parte degli strumenti digitali che elaborano i dati in vista di una loro conservazione, della loro analisi sotto diversi aspetti e della loro fruibilità sotto varie forme e con i limiti previsti dalle disposizioni normative. Contestualmente, la loro analisi contribuisce a restituire un quadro più chiaro della società europea nel suo complesso. Noi sappiamo, tuttavia, che i dati digitali, se non sono trattenuti

da vincoli normativi, sono nomadi e il loro viaggiare non può essere scevro da contaminazioni. In tutte le società del mondo, durante la pandemia causata dal Covid-19, ci sono state ricerche digitali e dati prodotti dalle stesse incentrate su questo tema. In termini sociali e culturali, ciò che emerge dall'analisi dell'enorme mole di dati rappresenta una miniera di informazioni e potrebbe rispecchiare un volto composito e interessante di come siamo oggi e di come la società attuale si evolve, non solo rispetto alla geopolitica sanitaria (Barbieri, 2023).

Per gestire ulteriori questioni complesse legate al mondo contemporaneo, per esempio quelle energetiche o climatiche, che stiamo vivendo in modo emergenziale, i ricercatori, attraverso l'impiego dell'intelligenza artificiale, prospettano una serie di strategie di intervento che sono il risultato di logiche analitiche e predittive sviluppate proprio con l'ausilio delle macchine. Pensiamo solo all'uso dei simulatori orientati a comprendere le dinamiche degli eventi naturali come i cambiamenti climatici o i terremoti, ma anche quelli orientati a migliorare le prestazioni di volo o di guida.

Un aspetto su tutti emerge chiaramente: anche la conoscenza della società odierna nella sua totalità è ormai mediata dal filtro della macchina.

Una componente decisiva del cambiamento culturale a cui stiamo assistendo è la presenza della riflessione etica, che interessa diversi ambiti in cui l'utilizzo dei dati digitali è presente ed entra in molte questioni di ordine legale, morale e filosofico, ingegneristico (Guerra, 2021; Budroni, 2021b). La creazione di codici etici da parte delle aziende o delle istituzioni che utilizzano l'intelligenza artificiale è diventata una questione estremamente attuale, poiché punta a orientare i comportamenti favorendo un clima di tutela e di fiducia reciproca tra i soggetti coinvolti (Guerra, 2023). Gestire gli aspetti tecnici, legali ed etici cercando di temperare diverse esigenze nel rispetto delle regole fa dunque parte di una buona governance. La Risoluzione europea del 20 ottobre 2020 e la successiva proposta normativa in discussione forniscono raccomandazioni affinché si stabilisca una cornice di riferimento etico e giuridico relativa all'intelligenza artificiale, alla robotica e alle tecnologie ad essa correlate. Essa spinge verso un approccio antropocentrico e antropogenico, sottolineando la priorità dell'intervento umano, che deve governare la realtà digitale (European Parliament Resolution, 2020). Al contrario, fuori dall'Europa, la ricerca di frontiera sembra essere più possibilista.

Come il linguaggio dei big data cambia la comunicazione e l'accesso alla conoscenza

6. I principi dei dati FAIR, sostenibilità dei dati, accesso aperto e comunicazione della scienza

Nell'era della rivoluzione digitale è evidente che l'essere umano non governa più autonomamente il processo di conoscenza, ma si deve affidare alle macchine. La spaventosa mole di dati che continuamente si autoproduce, senza una bussola con cui orientarsi al suo interno, anziché promuovere la conoscenza rischia di opprimerla e oscurarla. Per questo, a partire dal 2014, sono stati elaborati dei principi guida con l'intento di facilitare l'accesso ai dati e il loro riutilizzo. Con l'acronimo FAIR si accorpano questi concetti orientatori: *findable* (reperibile), *accessible* (accessibile), *interoperable* (interoperabile) e *reusable* (riusabile). I principi FAIR, se vengono applicati direttamente sui dati, fanno sì che gli umani e le macchine possano reperire questi dati attraverso degli identificatori persistenti. Una volta trovati, il loro grado di accessibilità viene definito attraverso sistemi di autorizzazione e autenticazione applicati ai dati o ai metadati; se essi saranno formattati con un linguaggio condiviso in modo da essere leggibili secondo degli standard internazionali, sarà possibile il loro riuso e la loro ricombinazione.

Il volume di dati digitali è così grande che la necessità di stabilire delle regole si estende anche alla loro conservazione e alla loro cancellazione. Per esempio, i dati conservati non sono necessariamente accessibili (ricordiamo la legge sulla privacy in base al Regolamento UE 2016/679) e possono essere cancellati (diritto all'oblio).

In Italia, il Piano Nazionale per la Scienza Aperta (PNSA), in attuazione al decreto ministeriale n. 268 del 28 febbraio 2022, recepisce le indicazioni europee e i pilastri su cui si fonda quello che viene definito come un vero e proprio approccio «al processo scientifico basato su collaborazione, condivisione aperta e tempestiva dei risultati, modalità di diffusione della conoscenza basate su tecnologie digitali in rete e metodi trasparenti di validazione e valutazione dei prodotti della ricerca» (PNSA, 2022). Da esso si evince che la collaborazione e la solidarietà fra gli scienziati trae vantaggio da un accesso gratuito e tempestivo ai risultati della ricerca, dalla possibilità di reperire facilmente in rete pubblicazioni scientifiche, ma anche dati e metadati che ne consentano il riuso gratuito senza barriere giuridiche o di sistema. L'accesso aperto si estende inoltre anche all'utilizzo delle infrastrutture digitali.

Durante la pandemia da Covid-19, le difficoltà di gestione hanno messo in luce come un accesso limitato alle pubblicazioni – che rappresentano i risultati della ricerca – o una parziale condivisione dei dati fra ricercatori di diversi

paesi abbia rallentato le possibilità di approntare soluzioni di contenimento o di studio del fenomeno. Ecco perché, si legge nel rapporto, l'Italia, recependo le indicazioni europee, si pone tra gli obiettivi futuri quello di fornire un accesso aperto immediato alle pubblicazioni scientifiche che vengono finanziate con fondi pubblici (cfr. Raccomandazione 790 della Commissione UE del 25 aprile 2018 sull'accesso e la conservazione della informazione scientifica), ma anche quello di incentivare le pubblicazioni non commerciali, cercando di mitigare gli oligopoli dei grandi gruppi editoriali che, con l'acquisizione gratuita dei diritti patrimoniali d'autore, controllano, specie in aree importanti della ricerca, la comunicazione scientifica senza pagare autori e revisori. A questo scopo, i rapporti con gli editori sono in corso di ridefinizione in molti paesi europei con l'obiettivo di ottenere l'accesso aperto (cosiddetto *gold*) immediato in sedi editoriali aperte e il deposito delle pubblicazioni in archivi aperti (cosiddetto *Green*), vincolato però, con un embargo temporale, alle regole di copyright (PNSA, 2022). Parimenti, quello a cui si punta è anche la creazione di infrastrutture e piattaforme comuni facilmente accessibili e fruibili ai ricercatori dei paesi europei e di altri paesi extraeuropei per abbattere le barriere di sistema. Oggi un video può essere facilmente caricato su YouTube perché l'accesso a questa piattaforma è aperto, mentre in molti atenei le politiche universitarie decidono se permettere o meno ai loro ricercatori di accedere ai dati dei risultati di ricerca, e decidono anche l'accesso ai sistemi stessi per poter caricare i video o gli audio. In questo spazio tra l'accesso diretto e il protezionismo dei dati si gioca la transizione verso la scienza aperta che sta tranquillamente avvenendo a livello mondiale (Budroni, 2021a).

Un'osservazione importante resta infine da fare riguardo al rapporto che i big data hanno rispetto allo sviluppo sostenibile fissato dalle Nazioni Unite e orientato a realizzare diversi obiettivi di sviluppo (sustainable developed goals). Riguardo a questo tema, un recente studio condotto in Cina (Liu *et al.*, 2023) ha evidenziato come le città costituiscano un valido modello globale per l'attuazione degli Obiettivi di Sviluppo Sostenibile (SDG). Nei paesi in via di sviluppo o sottosviluppati, le lacune di conoscenza nella valutazione quantitativa dei progressi, determinate dalla carenza di dati statistici rilevanti per molte città, soprattutto quelle piccole, possono essere colmate attraverso l'uso dei big data. Utilizzando big data open source per 254 città cinesi e confrontando i risultati con quelli ottenuti impiegando dati statistici, i ricercatori hanno confermato che i big data sono un'alternativa promettente per monitorare il progresso comples-

Come il linguaggio dei big data cambia la comunicazione e l'accesso alla conoscenza

sivo degli Obiettivi di Sviluppo Sostenibile delle città, comprese quelle prive di dati statistici rilevanti (83 città cinesi).

7. Conclusioni e prospettive future. Verso una filiera virtuosa della conoscenza

L'universo dei big data ha cambiato sostanzialmente l'approccio alla conoscenza e alla comunicazione scientifica, ma come si può stare al passo con una rivoluzione in costante accelerazione? L'iniziale distinzione tra il mondo digitale e quello reale è diventata sempre più sfumata, tanto da richiedere una maggiore consapevolezza in termini di competenze da parte di tutti i cittadini, che vivono in maniera immersiva, eppure in parte inconsapevole rispetto a questo mutamento di paradigma. Un fatto concreto è che il mezzo (la macchina) è diventato uno strumento imprescindibile nella produzione di conoscenza, l'interfaccia *sine qua non*.

Stando così le cose, la gestione dei dati digitali, oltre a richiedere un impegno globale sotto vari profili (etici, economici, culturali, sociali, legislativi), ha bisogno di una corretta e contestuale comunicazione di quello che sta avvenendo, di costruire una sorta di filiera virtuosa della conoscenza, perché il suo sviluppo sostenibile passa attraverso la formazione degli attori principali, ma anche attraverso una corretta informazione e trasmissione, che deve avvenire durante tutto il percorso di produzione del sapere. Una buona comunicazione contribuisce infatti a creare fiducia nella scienza e nel suo sistema organizzativo, promuove l'inclusione, soprattutto quando opera con trasparenza e in maniera aperta. Questa è una sfida in cui tutti gli attori coinvolti sono chiamati a fare la loro parte: i singoli paesi con le loro istituzioni scientifiche, i ricercatori e gli editori, le aziende private e le grosse società internazionali. Ovviamente a questa chiamata non sempre si può o si vuole rispondere in maniera omogenea. Un primo ostacolo è determinato dall'uso massivo di un nuovo linguaggio, un ibrido tra il linguaggio delle macchine e un inglese strumentale che cambia costantemente e si arricchisce di termini sempre nuovi con cui occorre familiarizzare. Chi non lo conosce e non lo parla è escluso dalla comunicazione. Ecco perché una corretta comunicazione della scienza si avvale anche di una educazione e una alfabetizzazione costanti.

Per quanto riguarda i dati digitali, le regole vengono da un'etica condivisa e da linee guida condivise a livello europeo e internazionale. Applicare i principi FAIR, rendere i dati non solo conservabili ma anche accessibili è quel passaggio decisivo che arricchisce il patrimonio culturale collettivo. Attraverso questo

accesso aperto, consentito ai ricercatori o anche ai cittadini esperti, a quelli che vengono chiamati *citizen scientists*, i dati vengono letti, analizzati, condivisi e interpretati per produrre nuova conoscenza, che è sempre più olistica e partecipata (Ciriaco *et al.*, 2023).

Lo sviluppo sostenibile, infine, si realizza anche tramite una politica di gestione dei dati quanto più responsabile e uniforme possibile, che vada a vantaggio della crescita globale e della lotta alle disparità. Questo si ottiene anche creando il cosiddetto *skill development* (sviluppo delle competenze), per formare in modo corretto i data stewards addetti alla creazione di metadati secondo i principi FAIR (Budroni *et al.*, 2023). Accanto a questo, una disseminazione dei risultati coerente con una *filiera aperta* e circolare nella produzione dei dati non può, a parere di chi scrive, che arricchire la comunicazione della scienza nel suo complesso e facilitare, come una buona bussola, l'orientamento dentro la complessità del vivere.

Riferimenti bibliografici

- Barbieri P. (2023), *Intelligenza artificiale e ingegno*, Book Time, Milano.
- Budroni P. (2021a), *About Open Science and Autonomy of Science*, «Data Intelligence», 3 (1), pp. 52-63, https://doi.org/10.1162/dint_a_00077.
- (2021b), *Scienza, conoscenza, identità reale e identità virtuale*, «La Filosofia futura», 16, <http://digital.casalini.it/5198669>.
- Budroni P., Fátima A., Sánchez G., de Mello Castro Giroletti J., Sánchez Solís B. (2023), *Developing a Pan-European Open Science Training Landscape. The Project Skills4EOSC and Selected Initiatives*, «ABI Technik», vol. 43, 2, pp. 68-77, <https://doi.org/10.1515/abitech-2023-0014>.
- Cancila D., Mazzanti S. (2009), *Dizionario enciclopedico di informatica inglese-italiano italiano-inglese*, Zanichelli, Bologna.
- Chawinga W.D., Zinn S. (2019), *Global perspectives of research data sharing: A systematic literature review*, «Library & Information Science Research», 41, 2, pp. 109-122, <https://doi.org/10.1016/j.lisr.2019.04.004>.
- Ciriaco S., Fantin M., Scrigner C., Faresi L., Furfaro G., Trainito E., Segarich M., Spoto M. (2023), *Aggiornamento della presenza di “Nudibranchi” nel golfo di Trieste – il valore della citizen science*, «Biologia Marina Mediterranea», 27 (1), pp. 125-128.
- Convegno e report dell'Osservatorio del Politecnico di Milano (2022), «Internet of Things: tra continuità e innovazione».

Come il linguaggio dei big data cambia la comunicazione e l'accesso alla conoscenza

- Del Marco V., Pezzini I. (2017) (a cura di) *Nella rete di Google. Pratiche, strategie e dispositivi del motore di ricerca che ha cambiato la nostra vita*, FrancoAngeli, Milano.
- European Parliament Resolution (2020), *European Parliament Resolution of 20 October 2020 with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics, and related technologies*, <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020IP0276>; [https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI\(2021\)698792](https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2021)698792).
- GDPR (2018), *General data protection regulation*, https://europa.eu/youreurope/business/dealing-with-customers/data-protection/data-protection-gdpr/index_it.htm.
- Guerra E. (2021), *Rem tene verba sequentur. Big data e questione etica*, «La Filosofia futura», 16, <http://digital.casalini.it/5198666>.
- (2023), *A Code of Ethics for Social Cooperative Robots*, in V. Petuya, G. Quaglia, T. Parikyan, G. Carbone (eds), *Proceedings of I4SDG Workshop 2023, I4SDG 2023, «Mechanisms and Machine Science»*, vol. 134. Springer, Cham, https://doi.org/10.1007/978-3-031-32439-0_44.
- Liu Y., Huang B., Guo H. et al. (2023) *A big data approach to assess progress towards Sustainable Development Goals for cities of varying sizes*, «Commun Earth Environ», 4, 66, <https://doi.org/10.1038/s43247-023-00730-8>.
- PNS (2022), *Piano nazionale per la scienza aperta*, https://www.mur.gov.it/sites/default/files/2022-06/Piano_Nazionale_per_la_Scienza_Aperta.pdf.
- Rolling Plan for ICT standardization, (2023), <https://joinup.ec.europa.eu/collection/rolling-plan-ict-standardisation/internet-things>.
- Xiao G. et al. (2020), *The virtual knowledge graph system ontop*, in *International Semantic Web Conference*, Springer International Publishing, Cham, pp. 259-277, https://doi.org/10.1007/978-3-030-62466-8_17.

