Biblioteche data-driven: il percorso del sistema bibliotecario del Comune di Milano

ELISA BELLISTRI
Responsabile dell'Unità Servizi Digitali e Automazione
dell'area Biblioteche del Comune di Milano
elisa.bellistri@comune.milano.it

LIÙ PALMIERI
Responsabile dell'Ufficio Innovazione e Sviluppo
dell'area Biblioteche del Comune di Milano
liu.palmieri@comune.milano.it

DOI: 10.3302/0392-8586-202206-019-1

Dov'è la vita che abbiamo perduto vivendo?
Dov'è la saggezza che abbiamo perso con la conoscenza?
Thomas Stearns Eliot, Cori da *La Rocca*

A gennaio 2021 il Sistema bibliotecario del Comune di Milano ha intrapreso il percorso per l'ideazione, redazione e implementazione di un piano strategico in grado di informare il suo operato nei prossimi anni. Il piano, che al momento della scrittura di questo articolo è alle sue battute finali di redazione, seleziona e focalizza una serie di assi strategici che sono stati perimetrati partendo innanzitutto dalle priorità che la città di Milano ha assunto su di sé e che in filigrana lasciano intravedere i macro-obiettivi dell'Agenda 2030: come biblioteche, abbiamo infatti avvertito forte l'esigenza di incardinare in modo più continuativo e pronunciato la nostra attività in un orizzonte strategico elaborato a livello cittadino, lavorando in modo solidale, congruo e pertinente, insieme a tutti gli attori che vi contribuiscono.
Il percorso per focalizzare il posizionamento strategico delle biblioteche in questo panorama ha visto due step principali: uno volto a recepire le sollecitazioni e le priorità che provenivano da soggetti interni (dipendenti) e l'altro volto a captare la posizione dei portatori di interesse su alcuni temi emersi da una disamina accurata dei due documenti strategici più ricchi dal punto di vista dell'articolazione e del campionamento che erano disponibili a livello cittadino:

- *Milano 2046 - Laboratorio per un futuro comune*:[1] indagine prospettica a lungo termine sul futuro promossa a partire da dicembre 2017 dalla Presidenza del Consiglio comunale di Milano che, attraverso contributi molto diversificati tra loro e afferenti a sfaccettati livelli della società (esperti, *stakeholders*, esponenti della società civile come università, associazioni, imprese) ha approfondito nel quadro del BES dimensioni rilevanti per la città; tra queste, la migrazione, il rapporto tra generazioni all'interno del ciclo di vita, la diffusione e l'accesso alla cultura, la qualità dell'ambiente e dei servizi, le principali variabili che determinano il benessere soggettivo e quello collettivo allo scopo di offrire agli organismi di governo della città delle plausibili direttrici d'azione che aumentassero la qualità urbana.
- *FareMilano*:[2] un percorso voluto a ottobre 2020 dal sindaco Sala per disegnare la Milano di domani tramite incontri di 37 tavoli di lavoro (curati da Fondazione Feltrinelli, Centro di Ricerca Green, Centro Studi PIM, Politecnico di Milano, Istituto per la Ricerca Sociale, Fondazione Veronesi e Fondazione Triulza) ai quali si sono seduti oltre 700 rappresentanti, che si sono espressi su sette temi: La metropoli dei quartieri; In transizione ambientale; Una città in salute; Smart & Working; Il bisogno di Milano; Nascere, crescere e vivere a Milano; Una città che crea, sa e forma.

Da queste due fasi, espletate attraverso questionari (per soggetti interni ed esterni) e focus group (solo per soggetti esterni) è scaturita l'ossatura del piano strategico,

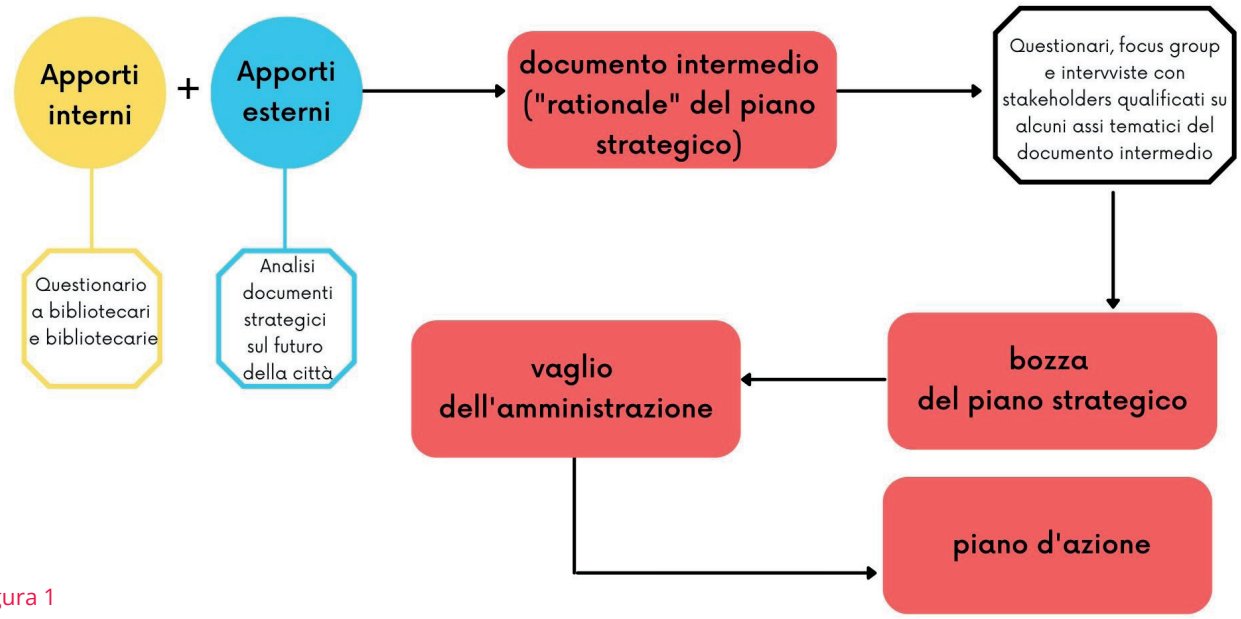


Figura 1

articolata in quattro assi principali (*Transizione ambientale*, *Transizione digitale*, *Città della prossimità*, *Promozione della lettura*) e in tre assi trasversali (*Accessibilità e inclusione*, *Partecipazione*, *Internazionalizzazione*) suscettibili di informare di sé tutte le azioni afferenti agli altri ambiti; un impegno a lungo termine che mira a ottimizzare ulteriormente l'efficacia degli assi *Città della prossimità*, *Accessibilità e inclusione* e *Partecipazione* coincide infine con l'opera di riqualificazione delle sedi bibliotecarie.
Si veda un prospetto generale in Figura 1.
È evidente però che il piano strategico è, oltre che una meta, anche un percorso, e che in quanto tale ingenera maggiore consapevolezza di ciò di cui c'è bisogno affinché il cambiamento non resti solo un "ottativo" o un auspicio di impatto. Ecco perché, strada facendo, stiamo approntando strumenti e strategie che attrezzino quel cambiamento in modo da renderlo, ancor prima che misurabile a valle, implementabile a monte.
Quando si articola, ad esempio, il tema della *Città della prossimità* prefigurando servizi bibliotecari sempre più diffusi anche grazie a partnership territoriali, capacità di lettura del territorio al di là dei confini amministrativi, possibilità di spiegare comportamenti degli utenti delle biblioteche e – *a fortiori* – dei potenziali utenti delle biblioteche al momento interpretabili solo con dati impressionistici, diventa evidente che occorrano specifici strumenti per fare delle biblioteche delle leve efficaci di prossimità.
Ed è proprio qui che il piano strategico trova un alleato indispensabile in un'analisi dei dati che non sia compilativa ma strategica e orientativa, esattamente come viene già fatto in altri settori dell'amministrazione pubblica che operano su scelte data-driven.

Le strategie data-driven dell'amministrazione comunale

La digitalizzazione della pubblica amministrazione ha tra gli obiettivi una maggiore rapidità nello scambio di informazioni, una maggiore efficienza nell'evadere le richieste del pubblico, una aumentata interoperabilità tra i diversi uffici. Non solo: la digitalizzazione dei processi sta creando un'enorme mole di dati che non sono solo un sottoprodotto delle normali attività amministrative, ma al contrario una fonte importante di nuovi processi di analisi, i cui risultati impattano sulla rapidità, l'efficienza e l'interoperabilità già descritte.
L'uso appropriato di questi dati permette di attingere a informazioni corrette e verificabili, ossia a una base di conoscenza oggettiva che agevola la capacità soggettiva di prendere decisioni.
L'amministrazione comunale di Milano da diversi anni persegue una strategia di decisioni data-driven, parallelamente alla digitalizzazione, e l'impatto sui processi è già visibile e quantificabile. L'incrocio tra i dati della direzione Commercio con quelli della direzione Urbanistica ha permesso, per esempio, nel corso dell'emergenza pandemica, di velocizzare l'iter di approvazione delle domande di occupazione suolo pubblico per la ristorazione da 90 a 15 giorni, così come l'analisi dei tempi di attesa agli sportelli dell'anagrafe ha migliorato la gestione delle diverse pratiche, riducendo i tempi di attesa da una media di 90 minuti nel 2018 a soli 20 minuti nel 2020.[3]
Questi sono solo alcuni esempi, volti a descrivere il contesto nel quale la digitalizzazione anche all'interno delle biblioteche diventa il prerequisito per iniziare un

percorso di analisi dati, con lo scopo di ottenere maggiore efficienza e penetrazione del pubblico.
Prima ancora di iniziare un lavoro di analisi, tuttavia, bisogna chiedersi se i dati a disposizione permettano di compiere un percorso decisionale. I dati grezzi sono infatti solo la base di una piramide, i cui livelli successivi rappresentano le operazioni compiute sui dati e i risultati ottenuti (Figura 2). Nel livello successivo ai dati grezzi si trovano i dati contestualizzati, che generano un'informazione, e sono le relazioni tra i dati che costituiscono il livello della conoscenza, per finire poi con l'individuare un pattern preferenziale tra le relazioni evidenziate, come vertice della piramide. Il vertice, nel modello di piramide della conoscenza elaborato da Russell Ackoff[4] viene definito come saggezza, ma nel nostro contesto lo potremmo tradurre più prosaicamente come *capacità di prendere delle decisioni*.

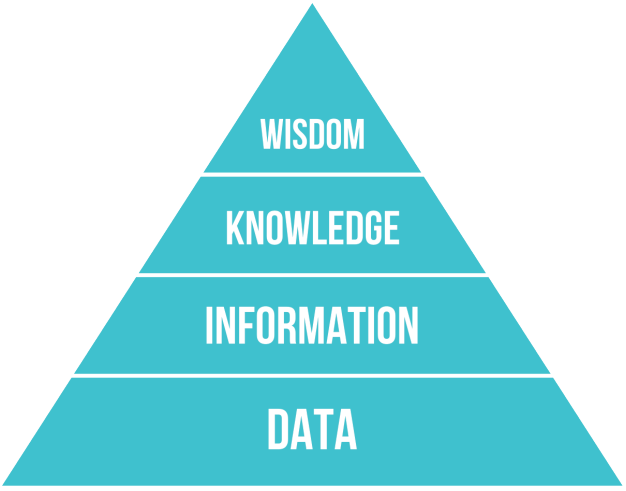


Figura 2

Ci siamo quindi interrogati sulle caratteristiche che devono avere i dati grezzi, cercando in prima istanza la corrispondenza con le 3V identificate da Doug Laney:[5] il volume, la velocità e la varietà: disponiamo di abbastanza dati? Sono dati in continua evoluzione e aggiornamento? Sono dati diversificati come tipologia e formato?
A titolo di esempio, alcuni valori relativi al volume dei dati dell'intero sistema bibliotecario a disposizione sono il numero di utenti attivi (160.000 nell'ultimo quinquennio, scelto per avere un orizzonte temporale stabile rispetto al fenomeno della pandemia) e il numero di titoli a catalogo, circa un milione. Queste due quantità si intersecano tra loro, generando la dinamicità del dato. Gli utenti attivi realizzano operazioni di prestito sui libri, un flusso che si attesta in media sui 3.000 prestiti al giorno. Allargando l'orizzonte delle attività della biblioteca al di là del solo prestito, diventa rilevante anche conoscere il numero di persone che transitano dalla biblioteca, per eventi, per laboratori, per studio. La raccolta dei dati si estende quindi all'uso di strumenti IoT (Internet of Things) come i sensori che registrano l'ingresso delle persone. Nella biblioteca Sormani, la biblioteca centrale di Milano, questi sensori già attivi registrano un flusso medio di circa 500 utenti ogni ora.
Da questa breve carrellata di dati disponibili emerge anche la loro varietà: dai dati di tipo bibliografico, a quelli anagrafici, ai dati geolocalizzati sino ai dati di affluenza provenienti da sensori IoT e dalle applicazioni di prenotazione per la partecipazione agli eventi.
L'uso di sensori intelligenti, quindi la produzione di dati che esula dall'intervento umano, aumenta enormemente la frequenza di acquisizione, concorrendo alla formazione di quel volume, quella massa critica, che separa i dati tradizionali dai *big data*. Lo scopo dei sensori è quello di intercettare comportamenti o parametri, trasformandoli in dati. Quindi l'ingresso all'interno della biblioteca, o di una parte delimitata di essa, viene registrata da un sensore di passaggio, ma analogamente anche un sensore rfid posizionato sul libro ne registra in tempo reale lo stato (in prestito, in restituzione) e la posizione (a scaffale, in deposito). Gli eventi registrati non sono più solo quelli inseriti manualmente da un operatore, ma costituiscono un flusso continuo in tempo reale, o quasi reale.
Una volta popolata la base della piramide, l'attenzione si sposta verso gli strumenti che permetto
no di raggiungere i livelli superiori: categorizzare i dati, trovare le relazioni, definire i percorsi preferenziali.
Per questo abbiamo sviluppato degli applicativi specifici, con il supporto di unità specializzate all'interno della direzione informatica del Comune: l'unità Sistemi informativi territoriali (SIT) per la gestione dei dati di tipo geografico e l'unità Analytics e Data Science per lo sviluppo di un applicativo di Business Intelligence (BI), che permetta di navigare i dati contenuti nel gestionale del patrimonio e dei prestiti. Questi applicativi, attualmente in fase di test, sono pensati per avere quotidianamente a disposizione degli strumenti che permettano di esplorare i dati, fuori dalla consueta struttura tabellare del database, ma in una veste grafica e con una esplicita struttura di relazione tra essi.

La dashboard di Business Intelligence

L'avvento dei Big Data ha reso indispensabile la progettazione di strumenti di tipo grafico per l'esplorazione dei dati, che possano dare con maggiore immediatezza un'immagine dei fenomeni sottostanti. Tutti abbiamo

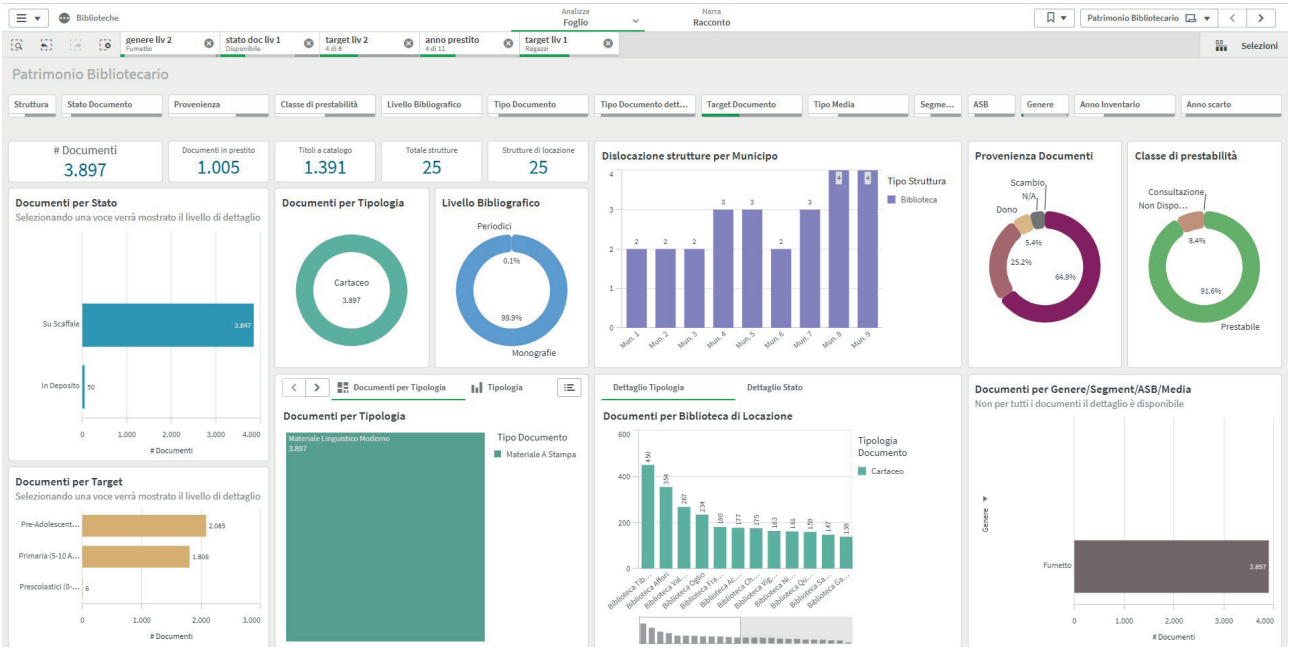


Figura 3

familiarizzato con le curve dell'epidemia di Covid-19, molte testate giornalistiche, enti governativi e altri organi di informazione hanno messo a disposizione del pubblico, sin dai primi mesi del 2020, dei cruscotti interattivi per visualizzare i dati dei contagi, dei ricoveri, dei decessi o delle vaccinazioni. L'immagine delle curve, con picchi crescenti rispetto ai mesi precedenti, le oscillazioni più o meno sincrone con altri paesi europei, e altre caratteristiche, hanno di sicuro avuto un impatto molto maggiore di quello che poteva dare una semplice lista di numeri.

Ma non è solo l'immediatezza visiva la caratteristica operativa di una dashboard di BI. Altri elementi distintivi, oltre ai grafici, sono i KPI e i filtri. Un KPI (Key Performance Indicator) è un numero che sintetizza un intero processo. Il numero totale degli utenti attivi in un lasso di tempo è la sintesi di tutti i comportamenti che gli utenti hanno avuto, da chi si è recato in biblioteca più volte in un mese, fagocitando libri con ardore, a chi in quel lasso di tempo è stato un utente distratto e occasionale. Così strutturato, il KPI non consente di restituire questa complessità. La proporzione percentuale tra gli utenti che realizzano un alto numero di prestiti e quelli che ne realizzano un basso numero, consegna invece maggiore profondità all'informazione sul comportamento di quel numero totale di utenti, pur mantenendo una prospettiva di sintesi. La scelta dei KPI deve quindi tenere conto simultaneamente sia della necessità di sintesi che di quella di profondità del dato, adeguandosi di volta in volta a seconda dell'informazione richiesta.

In modo speculare agiscono i filtri, che servono invece a disaggregare il dato, sia scendendo in profondità sul fenomeno sottostante, sia confrontando tra loro aspetti diversi dello stesso fenomeno. Il numero di prestiti effettuati è ancora una volta la sintesi grossolana di un aggregato di prestiti di diverse tipologie di materiale, ognuno con le sue specifiche peculiarità di assortimento e di pubblico preferenziale. Agendo con i filtri si modificano in modo dinamico i grafici e i KPI ampliando, come in un gioco di scatole cinesi, le informazioni presenti nella prima visualizzazione statica (Figura 3).

Illustriamo di seguito un esempio dell'uso dinamico dei filtri, per analizzare in dettaglio un fenomeno. Prendiamo per esempio il caso del solo prestito di fumetti, in particolare dei fumetti catalogati per target di età, come fumetti per bambini/adolescenti. La selezione di questa tipologia di materiale immediatamente modifica tutti i grafici e i KPI presenti sulla dashboard, mostrando lo stato al prestito, la provenienza, la quantità del solo materiale fumetti disponibile in ogni biblioteca. Anche gli utenti selezionati diventano i soli utenti che hanno preso in prestito dei fumetti.

In questo modo evidenziamo come le diverse biblioteche del Sistema bibliotecario di Milano si comportano rispetto all'attività di prestito di fumetti. Le prime due biblioteche, sia per quantità di materiale che per numero di prestiti, sono Tibaldi e Valvassori Peroni. Scompare Sormani dalla vetta della classifica (la biblioteca

centrale realizza il numero maggiore di prestiti totali), proprio perché ci siamo concentrati su un segmento e un target molto specifico.
Se ci focalizziamo sulle due sole biblioteche Tibaldi e Valvassori Peroni, possiamo confrontare la quantità di materiale disponibile con il numero di prestiti effettuati. Tibaldi possiede molto più materiale del segmento fumetti, mostrando una scelta di posizionamento per la biblioteca, ma la Valvassori Peroni, essendo una biblioteca molto più grande, associa la scelta di posizionamento, al secondo posto come collezione di fumetti, con il flusso di utenti molto elevato, diventando la biblioteca con maggior numero di prestiti del segmento fumetti (Figure 4 e 5).
Un altro aspetto che emerge, relativo questa volta alle caratteristiche degli utenti, è la fascia di età degli utenti che effettivamente registrano il prestito. Pur essendo materiale per bambini e ragazzi, la maggior parte dei prestiti sono intestati ad adulti. In questo caso emerge un comportamento degli utenti: i minori frequentano



Figura 4

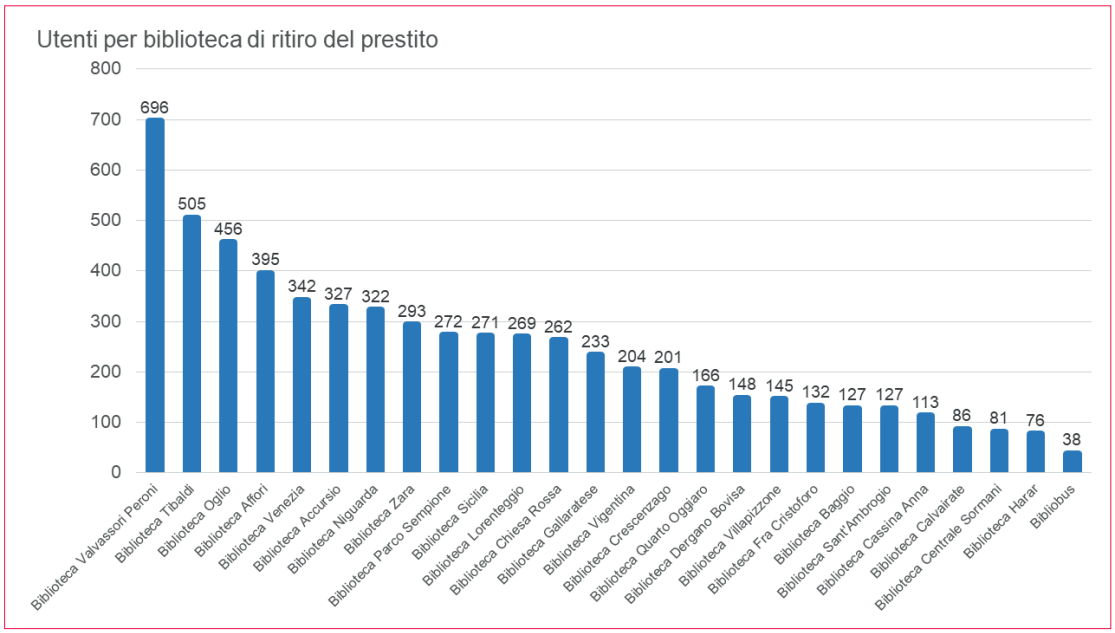


Figura 5

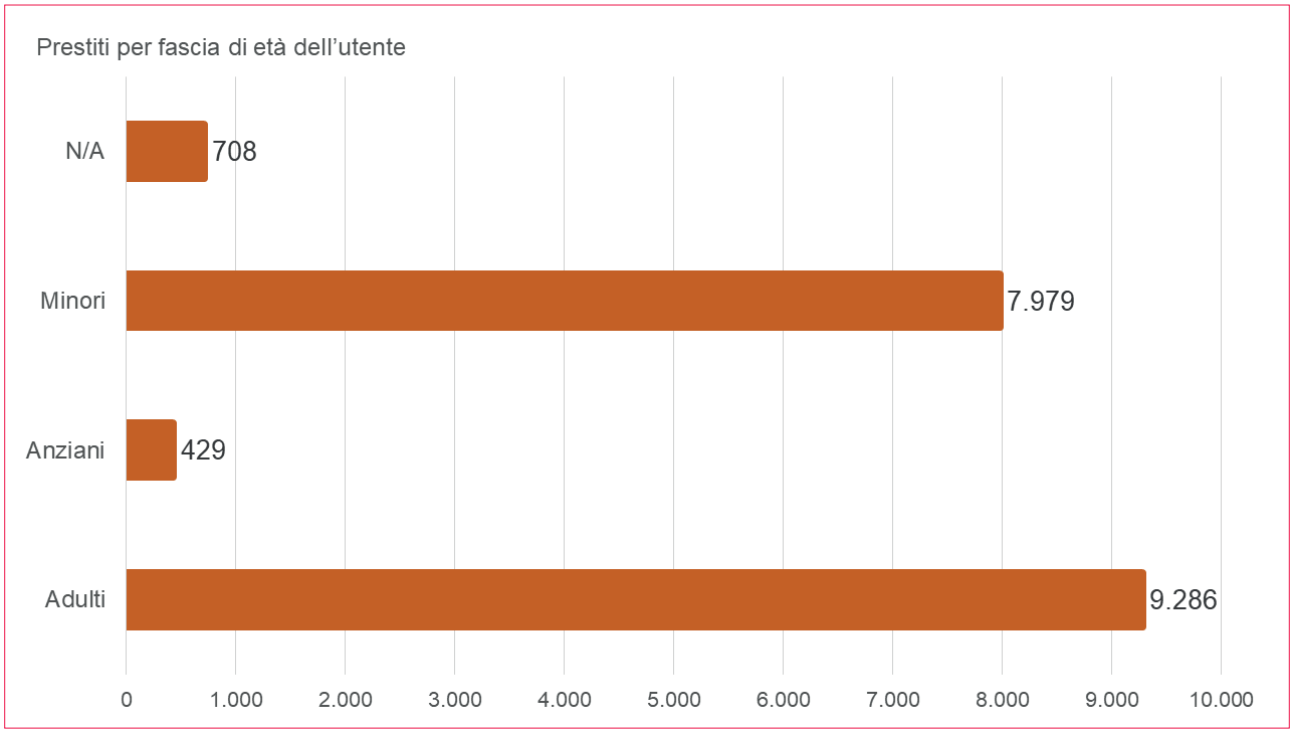


Figura 6

la biblioteca accompagnati da un adulto, che gestisce l'attività di prestito per conto del minore (Figura 6).
L'interattività dello strumento permette di proseguire l'analisi, partendo da questi primi risultati, con ulteriori selezioni e filtri, in modo flessibile, seguendo le domande che ogni risultato a sua volta pone. Questa modalità di esplorazione dinamica può portare a un set di risultati finali, ma anche gettare le basi per ulteriori modalità di analisi, di acquisizione dati, di pianificazione delle attività.

Il Geoportale

La possibilità di utilizzare dati di tipo spaziale aggiunge una ulteriore dimensione di profondità all'analisi dati. I dati spaziali legano le informazioni al territorio a cui appartengono, offrendo più livelli di lettura dei risultati. Il Geoportale del Comune di Milano, costruito su una piattaforma Esri, leader mondiale dell'analisi dei dati di tipo geografico, assolve al compito di mostrare i dati e la loro relazione con il territorio a cui appartengono. È quindi possibile visualizzare il numero di residenti in un determinato quartiere di Milano (NIL, nuclei di identità locale, che dividono la città in 88 settori secondo criteri demografici e strutturali), ma anche i residenti divisi per fascia di età, in modo da evidenziare il bacino di utenza di quella zona.
Una peculiarità degli strumenti di tipo geografico è quella di poter ospitare, sullo sfondo di una mappa, diversi altri strati di informazione (*layer*), attivabili o disattivabili, a seconda che si voglia avere la visualizzazione complessiva o separata per tipologia. Al layer della popolazione totale è quindi possibile aggiungere quello della popolazione iscritta alla biblioteca, ancora una volta con il dettaglio delle fasce di età, mettendo così in evidenza la correlazione tra i due insiemi di dati (Figura 7).
Anche la scelta dei sottoinsiemi geografici di riferimento, per esempio l'uso dei NIL precedentemente descritti, è un parametro modificabile che influisce sulla lettura del dato.
La rappresentazione dei dati di popolazione residente e popolazione iscritta alla biblioteca secondo il criterio dei quartieri, quindi un perimetro geografico predefinito, permette di estrarre informazioni facilmente intersecabili con altri dati resi disponibili dall'amministrazione comunale (per esempio il reddito medio pro capite per NIL in relazione al numero di iscritti alla biblioteca).
Contemporaneamente però, una rigida definizione dei quartieri non tiene conto dei movimenti reali nella città, che non si arrestano davanti a un virtuale confine amministrativo, ma si regolano secondo il criterio della prossimità. Nell'orizzonte della città della prossimità – uno degli assi del piano strategico di cui abbiamo scritto in apertura – la singola biblioteca è destinata a diventare un polo attrattivo per le persone che vivono nel suo intorno.

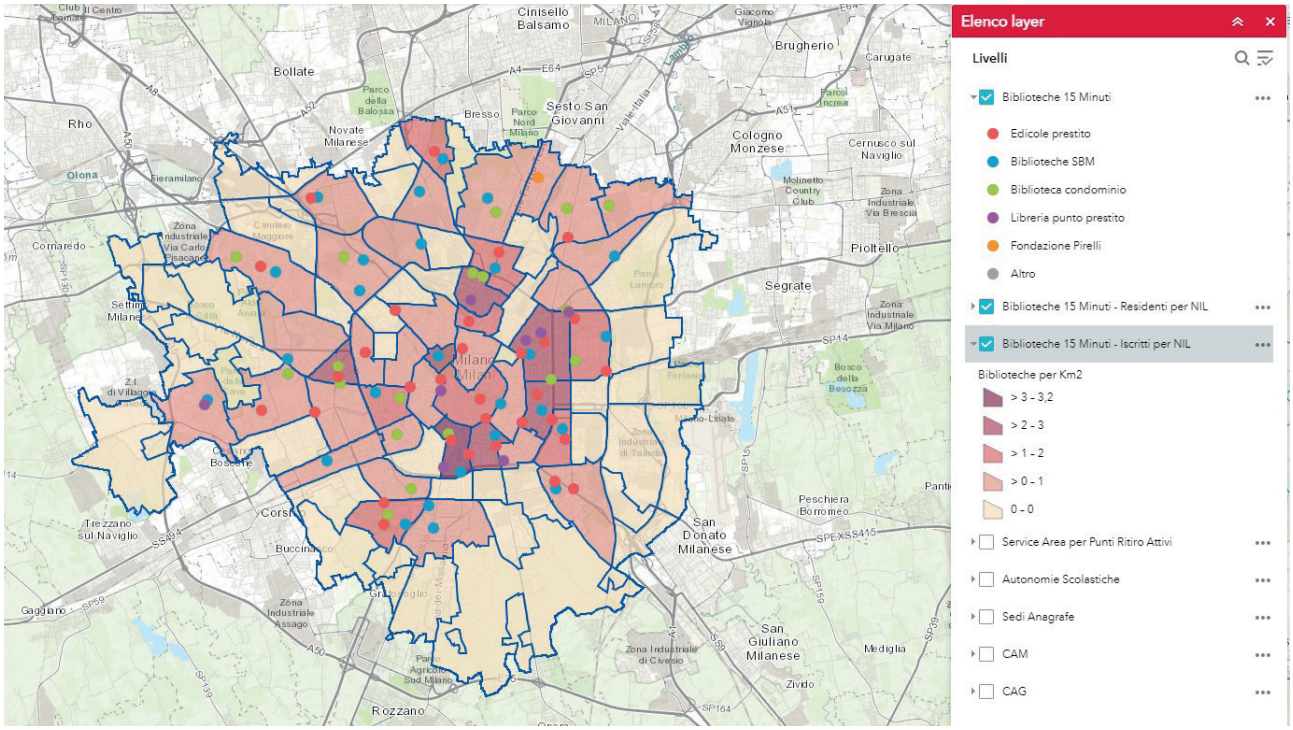


Figura 7 - Esempio di mappa geografica che riporta i confini degli 88 NIL, la densità di utenti iscritti alle biblioteche, la posizione delle biblioteche e altri punti di prestito.

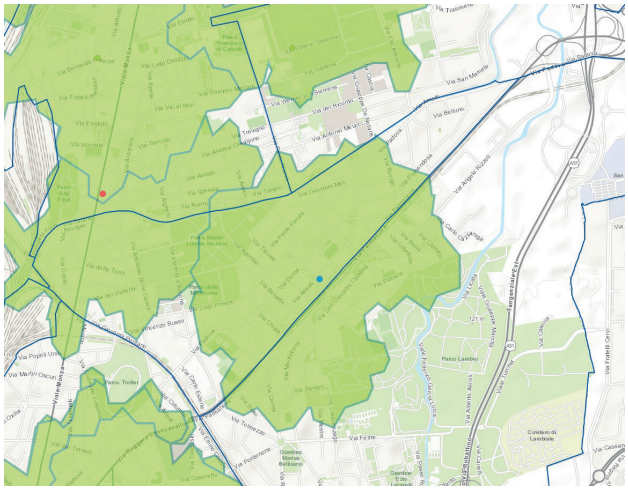


Figura 8 - Perimetro della service area di una biblioteca. Le linee blu sono i confini dei NIL

Lo strumento del Geoportale mette quindi a disposizione un algoritmo che, identificato un baricentro (per esempio una biblioteca), calcola tutti i possibili tragitti percorribili a piedi nell'arco di 15 minuti. Quello che si va a definire è un perimetro irregolare, che tiene conto delle barriere fisiche tra luoghi anche contigui (una ferrovia, un'area dismessa), e che racchiude al suo interno un insieme di residenti omogenei per quello che riguarda l'opportunità di raggiungere in tempi brevi la propria biblioteca di zona (Figura 8).

In questo cambio di prospettiva, in cui non si visualizza più un'area definita a priori, ma è la biblioteca stessa che in modo attivo definisce un'area, la domanda che viene più naturale porsi è: quanto la vicinanza della biblioteca influisce sulla frequentazione della biblioteca stessa?

Lo strumento del Geoportale consente quindi di estrarre i residenti ai civici contenuti all'interno di ciascun perimetro e di incrociarli con gli utenti attivi al prestito bibliotecario, aggregando i comportamenti di lettura con il criterio della prossimità.

Il risultato di questa analisi preliminare ci dice che la prossimità è un fattore che incentiva la frequentazione della biblioteca, ma da sola non è decisiva: se confrontato con il gruppo di utenti che vivono a più di 15 minuti di distanza, infatti, il 51,17% degli utenti di prestito cartaceo vive nelle vicinanze di una biblioteca, contro il 48,83% che frequenta la biblioteca pur non avendola nel raggio di 15 minuti.

Allo stesso tempo, analizzando il comportamento degli utenti rispetto alla singola biblioteca, si evidenzia che la sola prossimità non determina necessariamente comportamenti omogenei tra le diverse biblioteche.

Due casi limite sono quelli della biblioteca di Dergano e della biblioteca Sormani. Da un lato la biblioteca di

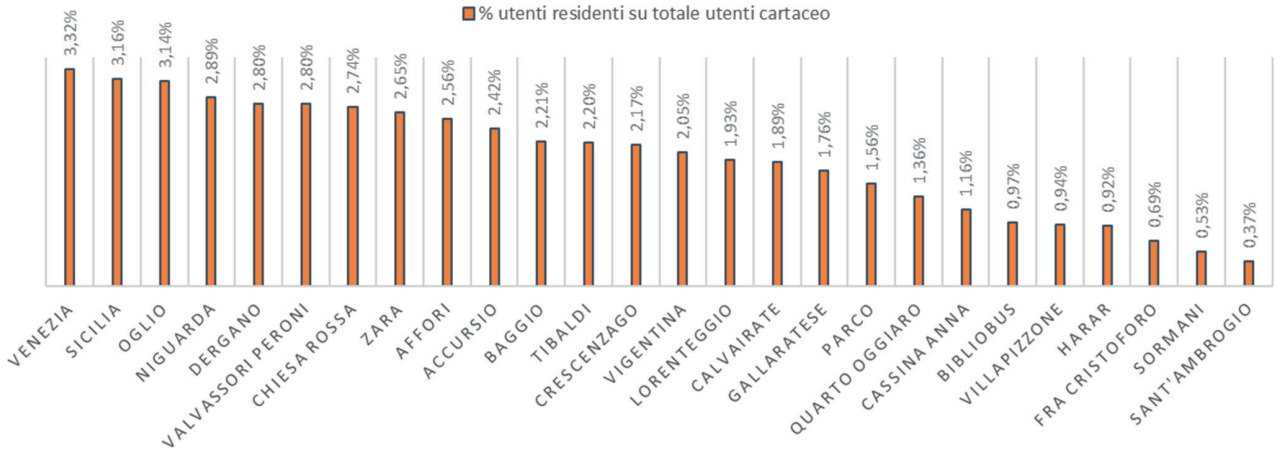


Figura 9 - Come si suddividono gli utenti del prestito cartaceo residenti in prossimità di una biblioteca, tra le diverse biblioteche

Dergano, un quartiere periferico di Milano con una forte identità locale, raccoglie in termini percentuali un altissimo numero di utenti tra i residenti, dall'altro la biblioteca Sormani si assesta agli ultimi posti come frequentazione percentuale da parte dei cittadini residenti nel suo intorno, sovrastati dal numero di utenti che provengono dall'esterno del suo perimetro di prossimità. Questo risultato, ovviamente non inatteso, definisce l'identità della biblioteca, più rivolta alla vita di quartiere in un caso, aperta a un pubblico vasto con esigenze differenziate e specialistiche nel caso della biblioteca centrale (Figura 9).

Questi primi risultati aprono a un elevato numero di domande ulteriori – Come cambia il valore della prossimità per fascia di età? Cosa spinge un utente a frequentare una biblioteca lontana da casa? Quanto rilevano gli spostamenti quotidiani nella frequentazione di una biblioteca? – la cui risposta permette di comprendere meglio il valore della biblioteca per gli utenti ed eventuali strategie per aumentarlo, e al contempo aggiunge elementi per la comprensione delle dinamiche in una città complessa come Milano, inserendosi così a pieno titolo nelle strategie data-driven della città.

Un fattore chiave: le competenze

Il lavoro che le biblioteche del Comune di Milano stanno impostando in quest'ambito è solo all'inizio, ma è già chiaro che lavorare con i dati rischia di essere una dispersione di risorse se l'estrazione, la lettura e l'interpretazione del dato non sono guidati da domande elaborate a livello strategico. Per questo motivo, è incalzante la necessità, per il futuro di biblioteche che si vogliano sempre più saldamente incardinate nell'orizzonte strategico della città in cui si trovano, di inserire negli staff competenze diversificate e inerenti ambiti tradizionalmente (ed erroneamente) considerati estranei al contesto bibliotecario. È stato grazie alla selezione e introiezione di una *data scientist* nel pool delle biblioteche di Milano che queste possono ora impostare articolate decisioni data-driven. Allo stesso modo, il cambiamento di cui il piano strategico è dispositivo principale va attrezzato con adeguati percorsi formativi di *capacity building* che dotino il personale delle biblioteche che vogliano divenire parte sempre più integrante della comunità di riferimento di competenze assimilabili a quelle, ad esempio, del *community manager*. L'Area biblioteche del Comune di Milano sta predisponendo a questo scopo un iter di formazione per tutto il personale bibliotecario, consapevole che il cambiamento vada supportato e attrezzato, non solo auspicato.

NOTE

[1] https://www.milano2046.it.

[2] https://www.faremilano.it.

[3] Intervista all'assessore Roberta Cocco al "Corriere Economia", 20 settembre 2021, https://tinyurl.com/2fbxjxs6.

[4] Russell Ackoff, *From Data to Wisdom*, "Journal of Applied Systems Analysis", 16 (1989), p. 3-9.

[5] Doug Laney, *3D Data Management: Controlling Data Volume, Velocity, and Variety*, Gartner, file No. 949, 2001, https://tinyurl.com/2p98yy6x.

ABSTRACT
Milan public library network has recently started to design its first strategic plan, which will provide a 4-year guidance for its 25 libraries. The strategic plan aims at reinforcing the role of Milan public libraries in the general strategic assets the whole city is pursuing in the future, such as ecological transition, digital transition, and the design of a 15-minutes city. In order to implement the decisions set by the strategic plan, libraries need to select, elaborate and use all the available data to take strategic decisions, as commonly done in a data-driven administration. This is why Milan public library is designing an approach to data that benefits from skills both of the library management and of data scientists.