

# Le collezioni messe a nudo

**DANILO DEANA**

Università degli studi di Milano

danilo.deana@unimi.it

**L**e biblioteche sono nate per raccogliere, conservare e mettere a disposizione dei loro lettori documenti prima manoscritti e poi a stampa. La diffusione delle risorse elettroniche ha però modificato la definizione di collezione. Oggi, infatti, con questo termine si indicano non solo i documenti, in qualsiasi formato e di qualsiasi tipo, che una biblioteca possiede, ma anche quelli a cui dà la possibilità di accedere.<sup>1</sup>

Questo cambiamento non ha fatto venir meno l'importanza della gestione delle collezioni. Con questo termine si indicano di solito tutte le attività legate a quest'ambito,<sup>2</sup> anche se alcuni preferiscono tenere distinto lo sviluppo dalla gestione. Peggy Johnson, autrice di una guida sull'argomento pubblicata dall'American Library Association, afferma che con "sviluppo delle collezioni" intende indicare il processo di sviluppo o di costruzione di una collezione in risposta alle priorità istituzionali, alle esigenze e agli interessi della comunità o dei singoli utenti che fanno capo alla biblioteca. Lo sviluppo delle collezioni comprende quindi la selezione, la definizione e il coordinamento della politica di selezione, la valutazione delle esigenze degli utenti e dei potenziali utenti, la gestione del bilancio, l'identificazione delle esigenze che la collezione deve soddisfare, la sensibilizzazione e il collegamento con la comunità e gli utenti, la pianificazione della condivisione delle risorse e, in alcuni casi, la revisione e la negoziazione dei contratti relativi alle risorse elettroniche. La gestione delle collezioni riguarda invece le decisioni relative a scarti, chiusura degli abbonamenti, spostamenti in magazzino, conservazione a lungo termine e le attività che informano tali decisioni: gli studi sull'uso della collezione stessa e il rapporto tra costi e benefici.<sup>3</sup>

La valutazione delle collezioni, sia che la si veda come

parte dello sviluppo o della gestione, è comunque un tassello fondamentale di queste attività.<sup>4</sup> La letteratura su questo tema, almeno quella in lingua italiana, ha però un approccio che si rifà ancora ai metodi messi a punto all'inizio degli anni Ottanta del secolo scorso dal Research Library Group, un consorzio di biblioteche fondato nel 1974 che successivamente è confluito nell'Online Computer Library Center (OCLC).<sup>5</sup> Allo scopo di valutare i diversi settori di cui è composta una collezione, infatti, si pensa sia ancora necessario stabilirne la copertura attribuendo un punteggio che può andare da 0 ("la biblioteca non possiede e non intende acquisire materiale relativo ad uno specifico settore") a 5 ("la biblioteca possiede collezioni che assicurano l'eshaustività in un determinato campo disciplinare").<sup>6</sup>

Per rendersi conto di quanto questo metodo sia anacronistico, basta esaminare gli studi su un periodo storico poco considerato fino quasi alla fine del secolo scorso: la tarda antichità. Andrea Giardina, in un suo articolo del 1999, parlava di "esplosione di tardoantico" riferendosi proprio al crescente numero di pubblicazioni relative ai secoli che hanno fatto da ponte tra l'antichità e il Medioevo.<sup>7</sup> Autori come Sidonio Apollinare – un alto funzionario dell'Impero Romano, vescovo di Alvernia dal 467 al 486 – che fino a qualche decennio fa erano considerati minori, oggi hanno siti loro dedicati in cui sono elencati centinaia di testi.<sup>8</sup>

Se dagli studi umanistici si passa a quelli scientifici, tecnici e medici (Scientific, Technical and Medical, STM), possiamo vedere come negli ultimi due secoli il numero degli articoli e quello dei periodici siano cresciuti rispettivamente del 3 e del 3,5 per cento l'anno. Nel 2018 le riviste attive erano circa 42.500 e gli articoli pubblicati oltre tre milioni. Il ritmo di crescita è

ulteriormente aumentato negli ultimi anni, passando al 4 per cento per gli articoli e al 5 per cento per le riviste, in conseguenza della crescita delle spese per la ricerca e lo sviluppo e del numero dei ricercatori, stimati tra i 7 e gli 8 milioni a seconda dei criteri utilizzati per rilevarli.<sup>9</sup> La stessa situazione la si ritrova anche in altri campi, tanto che ormai si parla comunemente di *information overload*.<sup>10</sup>

L'impossibilità di valutare la completezza dei diversi settori di una collezione non significa rinunciare a stabilire se essa sia adatta a soddisfare le esigenze dei suoi utenti, ma solo che per farlo è necessario dotarsi di nuovi strumenti. L'analisi dei dati, infatti, è ormai alla portata di tutti grazie al diffondersi di linguaggi di programmazione e librerie di funzioni *open source*.<sup>11</sup> Una valutazione di questo tipo, come vedremo, permette anche di pianificare una corretta gestione degli scaffali e di stabilire, nel caso sia possibile muoversi liberamente tra di essi, se il modo in cui i documenti sono disposti sia quello più adatto a favorire l'incontro dei lettori con i loro libri.<sup>12</sup>

Lo stesso OCLC, il centro in cui è confluito il Research Library Group, ha messo a punto WorldShare Collection Evaluation, uno strumento che permette a una biblioteca di confrontare la sua collezione con quella di altre biblioteche in modo automatico,<sup>13</sup> così da poter prendere decisioni basate su dati certi.<sup>14</sup>

## La Biblioteca di Filosofia dell'Università degli studi di Milano

Come caso di studio è stata utilizzata la collezione della Biblioteca di Filosofia dell'Università degli studi di Milano.<sup>15</sup> La nascita della biblioteca risale alla fine degli anni Cinquanta. Dopo il completamento del restauro dell'antico Ospedale Maggiore di Milano, avvenuto proprio in quel periodo,<sup>16</sup> gli istituti di Filosofia e di Storia della filosofia si trasferirono infatti da via della Passione in via Festa del Perdono, nell'ala dell'edificio attiguo al cortile Ghiacciaia. Negli stessi anni, una raccolta di qualche centinaio di volumi, fino ad allora conservati presso la Biblioteca delle

facoltà di Giurisprudenza, Lettere e Filosofia, fu trasportata in locali accanto a quelli in cui si trovavano gli istituti e costituì il primo nucleo della futura Biblioteca di Filosofia.<sup>17</sup>

Nel 1982 la nascita del Dipartimento di Filosofia, dalla confluenza dei due istituti, diede un nuovo impulso allo sviluppo della biblioteca. Il 9 settembre 1999 la biblioteca, su delibera del Senato accademico, divenne di settore autonomo. Progetti e investimenti – deliberati da un Consiglio di biblioteca in linea con quanto previsto dal Sistema bibliotecario di Ateneo – vennero gestiti dal direttore, un membro del personale tecnico-amministrativo appartenente all'area delle biblioteche. Proseguì così il cammino di crescita, rafforzato dall'autonomia gestionale e decisionale della struttura, che nel 2000 si è dotata di un *Regolamento* e nel 2002 di una *Carta dei servizi*. Nel 2017, a seguito di una riorganizzazione che ha coinvolto l'intera amministrazione dell'Università degli studi di Milano, la Biblioteca di Filosofia è entrata a far parte del Servizio bibliotecario di Ateneo, la cui responsabilità è affidata a una neonata direzione.

La collezione della Biblioteca di Filosofia è composta da 77.760 documenti (tutti i dati citati in questo articolo si riferiscono alla situazione al 31 dicembre 2018).

La raccolta comprende 70.969 monografie e 6.786 annate di periodico. Si tratta quasi esclusivamente di materiale a stampa (77.305 documenti su 77.755). I documenti sono divisi in sette sezioni, elencate nella

Sezione	Descrizione	Documenti
3L.	Sezione generale	63.774
3L.AB.	Fondo Antonio Banfi e Daria Malaguzzi Valeri	124
3L.DF.	Fondo Dino Formaggio	2.413
3L.DVD.	Documenti multimediali su supporto fisico	178
3L.ESAMI.	Testi d'esame	796
3L.GV.	Fondo Giovanni Vailati	2.618
3L.JUD.	Judaica	824
3L.PER.	Periodici	6.786
3L.VHS.	Documenti multimediali su supporto fisico	272

Tabella 1

Sezione	Collocazione	Descrizione	Documenti
3L.	A.	Opere di consultazione	2.886
3L.	ANT.	Libri antichi	1.110
3L.	CD.	Documenti multimediali su supporto fisico	62

Tabella 2

Sezione	Collocazione	Descrizione	Documenti
3L.	01.	Filosofia orientale	216
3L.	02.	Autori del VI e V secolo a.C.	1.122
3L.	03.	Autori del IV secolo a.C.	2.989
3L.	04.	Autori del III, II e I secolo a.C.	688
3L.	05.	Autori del I, II e III secolo	1.523
3L.	06.	Autori del IV, V e VI secolo	838
3L.	07.	Autori del VII, VIII e IX secolo	384
3L.	08.	Autori del X e XI secolo	419
3L.	09.	Autori del XII secolo	627
3L.	10.	Autori del XIII secolo	1.067
3L.	11.	Autori del XIV secolo	676
3L.	12.	Autori del XV secolo	689
3L.	13.	Autori del XVI secolo	1.531
3L.	14.	Autori della prima metà del XVII secolo	1.609
3L.	15.	Autori della seconda metà del XVII secolo	1.509
3L.	16.	Autori della prima metà del XVIII secolo	857
3L.	17.	Autori della seconda metà del XVIII secolo	2.911
3L.	18.	Autori della prima metà del XIX secolo	2.916
3L.	19.	Autori della seconda metà del XIX secolo	2.119
3L.	20b.	Autori di lingua italiana	4.407
3L.	20c.	Autori di lingua cinese	38
3L.	20d.	Autori di filosofia del linguaggio	1.357
3L.	20e.	Autori di estetica	1.339
3L.	20h.	Autori di lingua francese	2.830
3L.	20i.	Autori di lingua tedesca	4.245
3L.	20k.	Autori di scienza della comunicazione	507
3L.	20l.	Autori di lingua inglese	4.022
3L.	20m.	Autori di lingua russa	216
3L.	20n.	Autori di altre lingue	334
3L.	20o.	Autori di filosofia della scienza	1.345
3L.	20p.	Autori di fisica	563
3L.	20q.	Autori di scienze naturali	2.173
3L.	20r.	Autori di matematica	418
3L.	20s.	Autori di logica	3.196
3L.	20u.	Autori di filosofia della storia	714
3L.	20v.	Autori di intelligenza artificiale	355
3L.	20w.	Autori di bioetica	272
3L.	20z.	Autori di antropologia	550
3L.	COLL.	“Loeb Classical Library”	253
3L.	DP.	Fondo Mario Dal Pra	5.036
3L.	GDN.	Fondo Guido Davide Neri	826

Tabella 3

Tabella 1 (3L è la sigla attribuita alla Biblioteca di Filosofia all'interno del Servizio bibliotecario dell'Università degli studi di Milano).

La sezione 3L.ESAMI. è una sezione temporanea:

Laura Frigerio, l'attuale responsabile della Biblioteca, in un articolo pubblicato nel 2009 su “Knowledge Organization” –<sup>19</sup> era pensato in funzione dello scaffale aperto.

all'inizio di ogni anno accademico vi vengono infatti inseriti i testi adottati nei corsi tenuti dai docenti del Dipartimento. Dopo tre anni, i volumi sono riportati nella sezione originale, a meno che non continuino a essere tra quelli adottati.

La sezione generale è a sua volta divisa in due gruppi: i documenti esclusi dal prestito (opere di consultazione, materiali multimediali e fondi speciali) e quelli ammessi al prestito.<sup>18</sup>

I documenti esclusi dal prestito sono divisi in tre collocazioni (Tabella 2).

I documenti ammessi al prestito, molto più numerosi, sono divisi in 41 collocazioni (Tabella 3).

Nel seguito dell'articolo considereremo solo i documenti inseriti nelle prime 38 collocazioni della Tabella 3. Sono stati esclusi quelli presenti in COLL., una collocazione riservata alla collana “Loeb Classical Library” pubblicata dall'Università di Harvard, nel Fondo Mario Dal Pra e nel Fondo Guido Davide Neri. Queste collocazioni ospitano infatti libri donati alla biblioteca e come tali hanno caratteristiche diverse rispetto alle altre.

I documenti inseriti nelle 38 collocazioni prese in considerazione sono 53.571 monografie disposte in scaffali liberamente accessibili a studenti e docenti.

Lo schema di collocazione in uso presso la Biblioteca di Filosofia è stato messo a punto negli anni Sessanta del secolo scorso da Corrado Mangione e Maria Assunta del Torre, due docenti oggi scomparsi che allora insegnavano Storia della logica e Storia della storiografia filosofica.

Lo schema – descritto nei dettagli da

A ogni autore all'interno di una collocazione è attribuito un numero progressivo che caratterizza sia le edizioni delle sue opere sia gli studi a lui dedicati. All'edizione critica dell'*Apologia di Socrate* curata da Michael C. Stokes, ad esempio, è stata data la segnatura 3L.03.T.001.100, mentre lo studio sulla filosofia di Platone di Franz von Kutschera riporta sul dorso un'etichetta con la scritta 3L.03.S.001.117. Le due segnature indicano che l'*Apologia di Socrate* è il centesimo (100) testo (T) di Platone (001), un autore del IV secolo a.C. (03), acquisito dalla biblioteca. Quello di von Kutschera, invece, è il centodiciassettesimo (117) studio (S) sull'autore. In questo modo i docenti e gli studenti che frequentano la biblioteca possono individuare immediatamente le ultime pubblicazioni di un autore o gli studi più recenti a lui dedicati.

La distribuzione delle monografie all'interno delle prime 19 collocazioni, quelle relative alla filosofia orientale e ai secoli dal VI a.C. al XIX, è molto variabile, così come il rapporto tra testi e studi (Grafico 1).

A farla da padrone tra le collocazioni relative agli autori del XX secolo sono quelle che raggruppano gli autori stessi sulla base della lingua piuttosto che quelle tematiche (Grafico 2). Unica eccezione, la collocazione riservata agli autori che si sono occupati o si occupano di logica. Già a partire dagli anni Novanta del secolo scorso, infatti, insegnavano presso il Dipartimento di Filosofia tre docenti che si occupavano di questo settore: Edoardo Ballo (Logica 2), Corrado Mangione (Logica 1) e Silvio Bozzi (Storia della logica). Collaborava poi con la biblioteca Miriam Franchella, che oggi insegna Istituzioni di logica presso il Dipartimento di Studi storici.

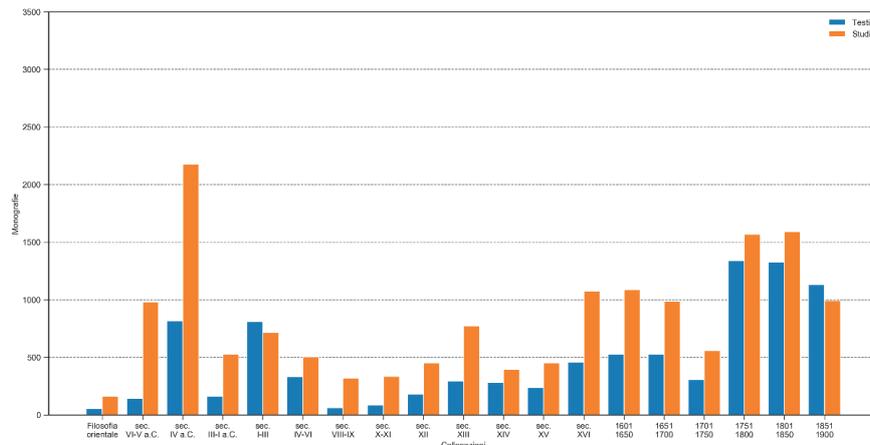


Grafico 1

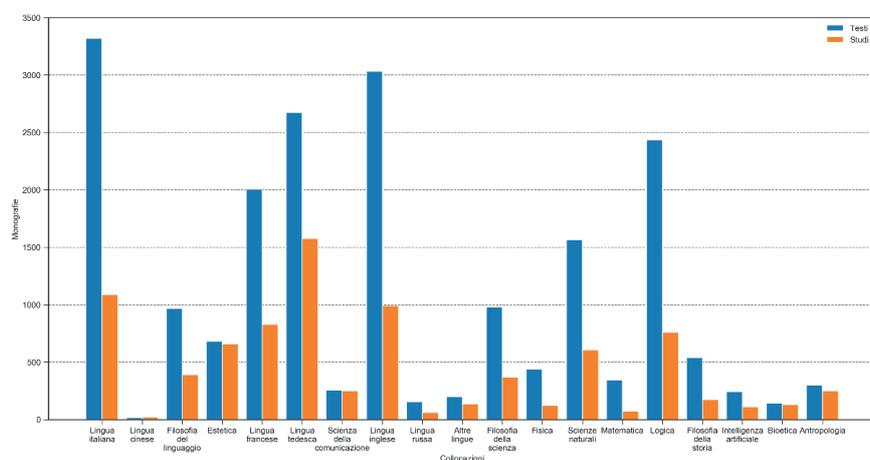


Grafico 2

## Obiettivi dell'indagine

I bibliotecari che lavorano presso la Biblioteca di Filosofia sono consapevoli che la loro collezione si è sviluppata in maniera non omogenea. Alcuni settori sono in crescita vertiginosa, mentre altri sono fermi da anni. Queste diverse velocità dipendono in parte dalla mancanza di una coerente politica di sviluppo. In assenza di una carta delle collezioni, il compito di individuare i volumi da acquisire è stato infatti lasciato ai docenti e agli studenti, che sono guidati prevalentemente dai loro interessi di ricerca o di studio del momento.

Per affrontare questa situazione, non sono sufficienti i dati raccolti annualmente sul numero delle nuove acquisizioni o sui prestiti, ma è necessario disporre di elaborazioni statistiche che siano in grado di correlare i dati tra di loro e di prendere in considerazione

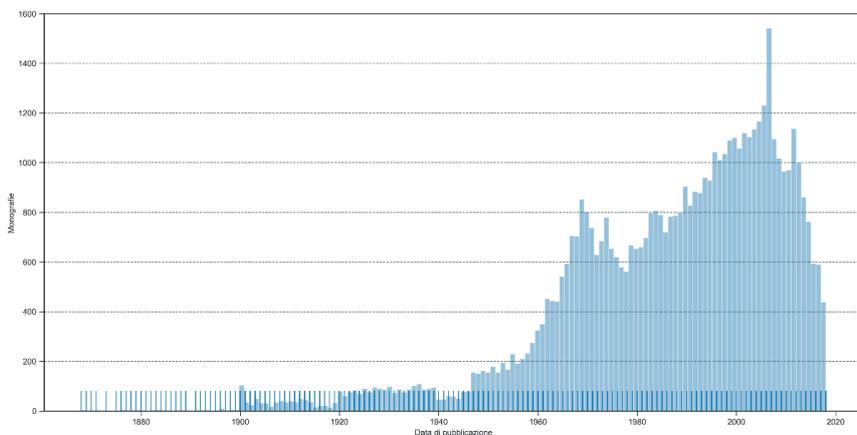


Grafico 3

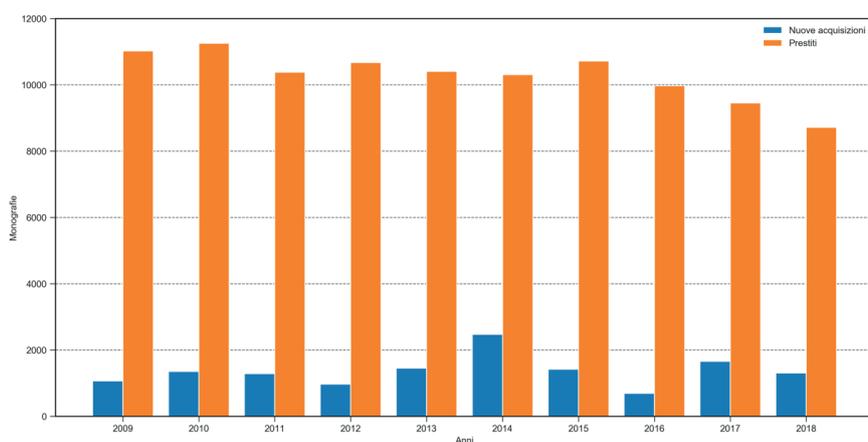


Grafico 4

periodi più lunghi. L'obiettivo che ci si è posti nell'avviare un'analisi di questo tipo è stato quello di individuare i punti critici della collezione: le sezioni e i numeri di catena nei quali, per usare termini mutuati dall'economia, l'offerta non incontra la domanda o la disposizione dei volumi non aiuta più gli utenti a muoversi tra gli scaffali.

## Sviluppo e uso della collezione

Il Grafico 3 rappresenta la distribuzione per data di pubblicazione delle monografie inserite nelle collocazioni ammesse al prestito della sezione generale e consente di seguire lo sviluppo della collezione dal 1859, l'anno della fondazione dell'Accademia scientifico-letteraria, fino al 2018. L'Accademia scientifico-letteraria costituì nel 1923 il primo nucleo dell'Università degli studi di Milano cui portò in dote il

suo patrimonio bibliografico. Se insieme all'incremento della collezione se ne esamina anche l'uso (Grafico 4), si può vedere come nel corso dell'ultimo decennio il numero delle nuove acquisizioni sia andato crescendo (il calo avvenuto nel 2016 è la conseguenza del fallimento di un fornitore, che ha comportato la mancata consegna degli ordini in corso), mentre quello dei prestiti è diminuito.

Uno degli indicatori più utilizzati per misurare la vitalità di una collezione è l'indice di circolazione.<sup>20</sup> Se nel calcolarlo si seguono le indicazioni dell'International Federation of Library Associations and Institutions, indicazioni che per le biblioteche accademiche prevedono si tenga conto solo di quanto acquisito negli ultimi cinque anni,<sup>21</sup> l'indice risulta in calo, dato che è passato dall'1,78 del 2009 (11.019 prestiti a fronte di 6.165 acquisizioni nel periodo 2005-2009) all'1,15 del 2018 (8.718

prestiti a fronte di 7.451 acquisizioni tra il 2014 e il 2018). Va però detto che nel 2014 sono stati catalogati 1.083 doni, contro una media negli altri anni di soli 150. Inoltre, nel 2009 è stata inaugurata la sezione 3L.ESAMI, di cui abbiamo parlato in precedenza e che non è stata presa in considerazione dato il suo carattere temporaneo. Nel 2018 le monografie all'interno di 3L.ESAMI, sono state prestate 1.622 volte contro le 424 del 2009. Se non si considera il picco dei doni del 2014 e si aggiungono i prestiti della sezione 3L.ESAMI, la differenza risulta meno accentuata: l'indice di circolazione passerebbe infatti da 1,85 a 1,58. Un'indagine condotta nel 2004 in 13 biblioteche olandesi aveva dato risultati compresi tra l'1,02 e il 4,89.<sup>22</sup> Nell'elenco dei dati che le biblioteche dovrebbero raccogliere compare il numero dei prestiti, ma non quello dei documenti interessati.<sup>23</sup> Nel nostro caso, i 102.288 prestiti verificatisi nei dieci anni considerati hanno riguardato 16.951 monografie. Solo 6.684 mo-

nografie sono state prestate una volta (Grafico 5). La vincitrice di questa particolare classifica è la seconda copia di *Filosofia del linguaggio* di Paolo Casalegno che è stata data in prestito 88 volte.<sup>24</sup> La prima copia – che fa attualmente parte della sezione 3L.ESAMI. – è stata data in prestito 160 volte.

Un altro dato di cui non è prevista la raccolta è la data di pubblicazione delle monografie prestate (Grafico 6). Nel caso della Biblioteca di Filosofia, durante il periodo che va dal 2009 al 2018 sono stati dati in prestito volumi la cui pubblicazione risale in media alla metà degli anni Novanta e che, con tutta probabilità, non sono più presenti nei cataloghi degli editori. È la conferma, se mai ce ne fosse bisogno, della fondamentale funzione di conservazione e messa a disposizione che hanno le biblioteche nel campo degli studi umanistici.

## Valutazione dei settori

Per analizzare nel dettaglio la collezione della Biblioteca di Filosofia è necessario passare dal livello delle collocazioni prese in considerazione, che come abbiamo detto sono 38 in tutto, a quello dei numeri di catena assegnati ai singoli autori, che sono 9.185.

Il Grafico 7 (per cui è stato necessario utilizzare una scala logaritmica) mostra come i numeri di catena che contengono fino a 10 monografie (le prime due barre dell'istogramma) siano 8.610, oltre il 95 per cento del totale. I numeri di catena con più di 50 monografie sono 158, meno del 2 per cento.<sup>25</sup>

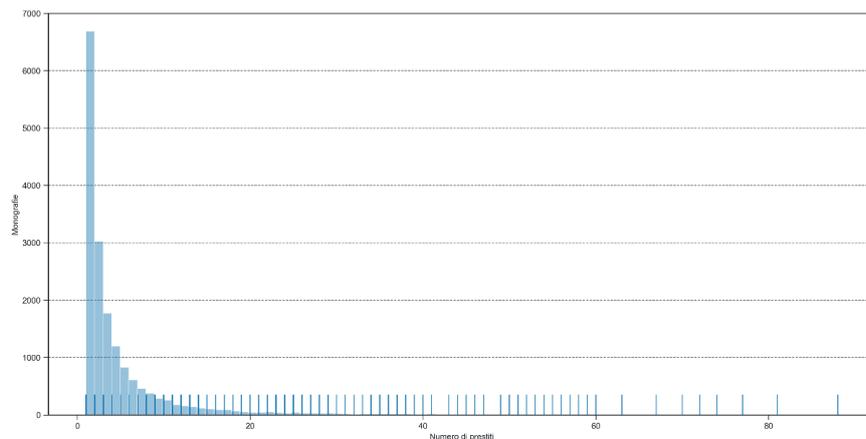


Grafico 5

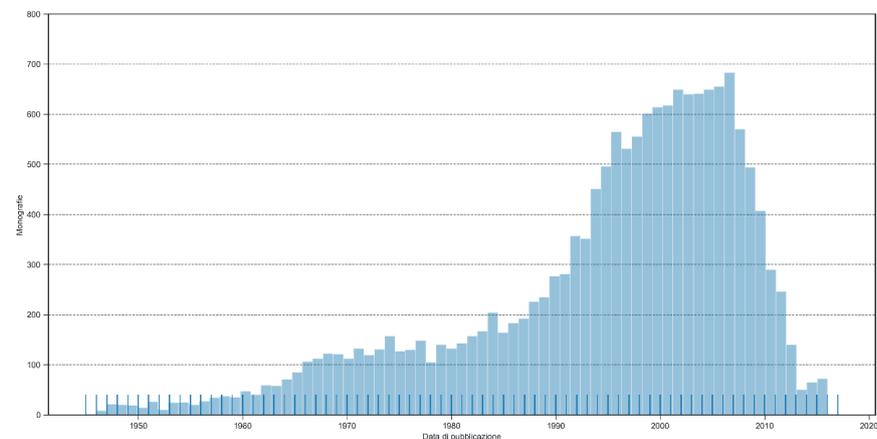


Grafico 6

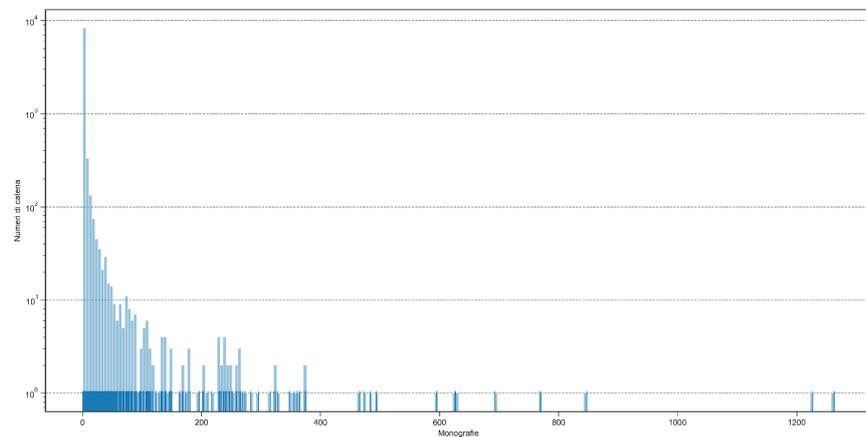


Grafico 7

Non era questo lo scopo per cui è stato progettato lo schema di collocazione. L'aumento esponenziale del numero delle pubblicazioni italiane e straniere ha però avuto l'effetto di moltiplicare i numeri di catena

nelle collocazioni del XX secolo. Gli autori che si occupano di argomenti non presenti nelle collocazioni tematiche finiscono poi in quelle basate sulla lingua, con il risultato di complicare ulteriormente le cose. Chi si muovesse lungo gli scaffali della collocazione dedicata agli autori di lingua inglese, ad esempio, passerebbe nell'arco di un solo ripiano da un saggio sulla religione di Samuel Martin Thomson, un filosofo americano, alla *Teoria generale dell'occupazione, dell'interesse e della moneta* di John Maynard Keynes, senza neppure la possibilità di vedere immediatamente quanto è stato pubblicato di recente negli Stati Uniti o in Gran Bretagna. Questo perché nello schema di collocazione per gli autori del XX secolo non sono state previste sezioni dedicate alla religione o all'economia.

La disomogeneità delle collocazioni basate sulle lingue può essere misurata anche analizzando i soggetti attribuiti alle monografie che ne fanno parte. Nella collocazione dedicata agli autori di lingua francese, ad esempio, sono state soggettate 1.633 monografie su 2.830 (le monografie soggettate sono solo quelle acquisite dopo il 1988, l'anno in cui è stata automatizzata la catalogazione; il recupero del pregresso, avviato nel 1991 a livello di Ateneo non "con libro in mano" ma "da scheda", non prevedeva infatti l'inserimento dei soggetti presenti sulle schede). I soggetti attribuiti sono stati 2.381, poco più di uno per ogni monografia. Per crearli si sono utilizzati 1.104 descrittori (le diverse parti che compongono un soggetto), con un numero di occorrenze pari a 4.194. Quest'ultimo dato mette in rilievo come si tratti di volumi eterogenei, dato che uno stesso descrittore è stato utilizzato meno di 4 volte. Per fare un paragone, nella collocazione 03 dedicata agli autori del IV secolo a.C. sono state soggettate 1.263 monografie su 2.989. I soggetti attribuiti sono stati 2.063 e per crearli si sono utilizzati 665 descrittori. In media ogni descrittore è stato utilizzato quasi 7 volte (in tutto le occorrenze sono 3.247).

Tornando ai numeri di catena, possiamo dire che analizzarne 9.185, a ognuno dei quali possono essere associate numerose caratteristiche (numero dei documenti, numero delle nuove acquisizioni negli ultimi dieci anni, numero dei prestiti negli ultimi dieci anni e così via), è un compito molto complicato. Non potendo esaminare tutti i dati nel dettaglio, si rischia di essere condizionati dai propri pregiudizi che por-

tano a privilegiare alcuni aspetti rispetto ad altri. È vero che, così come non ci sono pasti gratis, anche ogni algoritmo ha il suo prezzo,<sup>26</sup> ma se si scelgono con cura le caratteristiche da considerare, utilizzando procedure di *unsupervised learning* si può arrivare a conclusioni estremamente interessanti che possono guidare le azioni future.<sup>27</sup>

L'apprendimento non supervisionato è una delle tre tipologie in cui viene solitamente diviso l'apprendimento automatico (*machine learning*). Le altre due sono l'apprendimento supervisionato (*supervised learning*) e l'apprendimento per rinforzo (*reinforcement learning*).

L'*unsupervised learning* costituisce la base dell'apprendimento automatico. Secondo l'analogia di Yann LeCun, una delle maggiori autorità in questo settore, se l'intelligenza è una torta, la maggior parte della torta è l'apprendimento non supervisionato, la glassa sulla torta è l'apprendimento supervisionato e la ciliegina sulla torta è l'apprendimento per rinforzo.<sup>28</sup>

Le procedure di apprendimento non supervisionato sono particolarmente utili quando non si è capaci di stabilire preliminarmente come classificare gli oggetti presi in considerazione sulla base dei dati di cui si dispone. Esse, infatti, sono in grado di riunire oggetti che appartengono a uno stesso insieme (i *data points*) in gruppi omogenei chiamati *cluster*.<sup>29</sup> Nel nostro caso i *cluster* sono stati individuati attraverso l'algoritmo k-mean messo a punto da Stuart Lloyd<sup>30</sup> e Edward W. Forgy<sup>31</sup> tra gli anni Cinquanta e gli anni Sessanta del secolo scorso.

L'obiettivo dell'algoritmo è quello di ridurre al minimo la varianza totale all'interno dei *cluster*, identificati mediante un centroide o punto medio. L'algoritmo segue una procedura iterativa: inizialmente crea un certo numero k di cluster e assegna a ciascuno di essi i *data points* casualmente o usando alcune informazioni euristiche. Quindi calcola la distanza euclidea tra i *data points* e il centroide di ogni *cluster*. Se necessario, cambia la posizione del centroide o assegna un *data point* a un altro *cluster* per diminuire la distanza media dei *data points* dal centroide del *cluster* di cui fanno parte. La procedura viene ripetuta finché non viene raggiunto un risultato che non può essere ulteriormente migliorato.<sup>32</sup>

Grazie alla disponibilità di librerie *open source* molto ben documentate,<sup>33</sup> questo e altri algoritmi di *unsupervised learning* possono essere utilizzati anche da

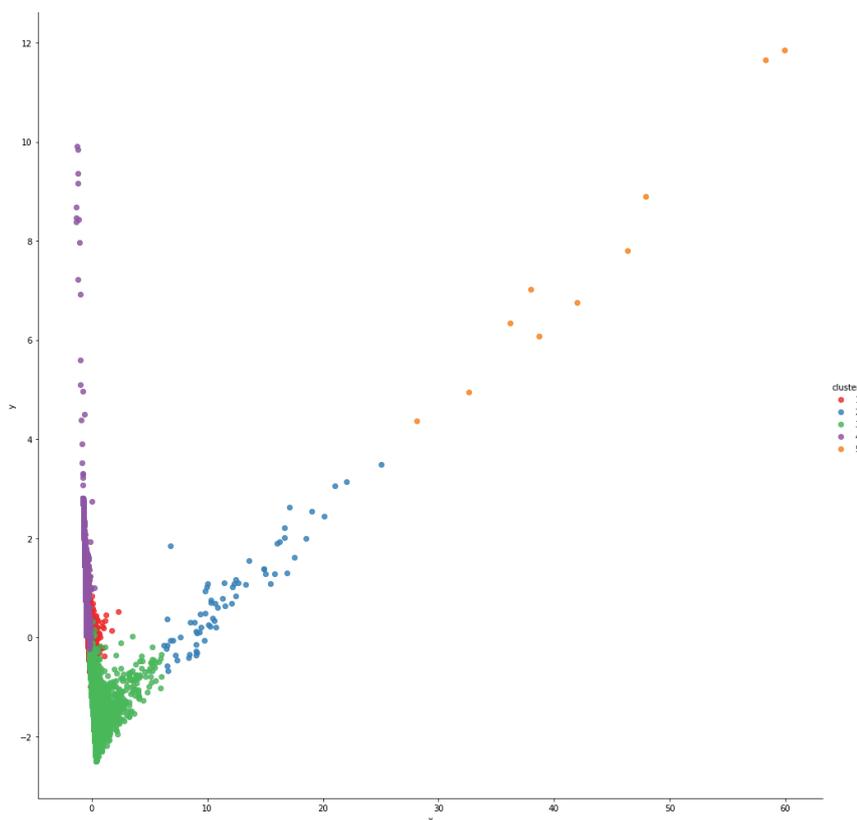


Grafico 8

non specialisti. Nel nostro caso, i numeri di catena in cui sono inserite le monografie ammesse al prestito della serie generale sono stati raggruppati in cinque *cluster* che è stato possibile rappresentare graficamente basandosi sulle loro caratteristiche principali (Grafico 8).

Nella Tabella 4 sono elencate alcune delle caratteristiche utilizzate per raggruppare i numeri di catena. La collezione presa in esame, come detto in precedenza, è composta da 53.571 documenti suddivisi in 9.185 numeri di catena. Nel decennio considerato le nuove acquisizioni sono state 13.705 a fronte di 102.885 prestiti.

Per interpretare il grafico è possibile utilizzare uno strumento messo a punto negli anni Settanta del secolo scorso, quando le grandi aziende statunitensi si trovarono nella necessità di valutare le diverse attività che venivano svolte al loro interno per essere in grado di allocare al meglio le risorse di cui disponevano. Il Boston Consulting Group, fondato nel 1963 da Bruce Henderson, mise allora a punto una matrice composta da due righe e da due colonne nella quale le attività erano classificate sulla base della loro quota di mercato e del tasso di crescita del mercato stesso.<sup>34</sup>

Le stelle (*star*) operano all'interno di un mercato caratterizzato da un alto tasso di crescita e ne possiedono una quota importante. Si tratta quindi di prodotti o servizi da curare in modo particolare e di cui è necessario programmare lo sviluppo. I punti di domanda (*question mark*) operano anch'essi all'interno di un mercato con un alto tasso di crescita, ma la loro quota è bassa. Essi assorbono risorse in conseguenza della crescita del mercato e della necessità di guadagnare posizioni rispetto ai concorrenti. In questi casi è necessario scegliere se sostenere l'attività o disinvestire. Le mucche da mungere (*cash cow*) possiedono una grande quota di un mercato caratterizzato da un basso tasso di crescita. Esse generano entrate elevate a fronte di bassi investimenti. I cani (*dog*), infine, sono attività con una esigua quota all'interno di un mercato con un basso tasso di sviluppo. Essi non assicu-

rano profitti soddisfacenti o addirittura generano delle perdite. L'unica strategia in questo caso è disinvestire.

Utilizzando i termini della Matrice BCG possiamo dire che il quarto *cluster*, 3.759 numeri di catena per un totale di 5.160 documenti, è quello dei cani: nessuna nuova acquisizione e nessun prestito negli ultimi dieci anni.

Il primo e il terzo *cluster* comprendono rispettivamente 18.897 e 6.132 documenti, quasi la metà del totale. Essi rappresentano i punti di domanda su cui si deve decidere se investire o meno. Il numero di prestiti è in entrambi i casi il doppio dei documenti, ma è andato diminuendo con gli anni, come dimostra il dato relativo all'ultima transazione. Le nuove acquisizioni poi non ci sono o sono inferiori alla media.

Il secondo *cluster* sono le mucche da mungere: 70 numeri di catena che raggruppano 15.131 documenti. Questi numeri di catena, infatti, producono oltre il 30 per cento dei prestiti (30.905 su 102.885) e assorbono una percentuale quasi corrispondente delle nuove acquisizioni (3.556 su 13.705).

Il quinto *cluster*, infine, è quello delle stelle: 11 numeri di catena che raggruppano 7.960 documenti. Pur comprendendo solo il 14 per cento dei documenti to-

Cluster	Numeri di catena	Documenti	Acquisizioni 2009-2018	Prestiti 2009-2018	Anni dall'ultima acquisizione	Anni dall'ultimo prestito
1	2.418	2,5	0,0	4,8	20,0	3,7
2	70	216,2	50,8	441,5	0,6	0,0
3	2.927	6,5	2,5	14,7	3,5	7,0
4	3.759	1,4	0,0	0,0	19,7	19,9
5	11	724,5	256,0	1569,8	0,6	0,0

Tabella 4

Sezione	Collocazione	Intestazione	Documenti	Acquisizioni 2009-2018	Prestiti 2009-2018
03	001	Plato	1.227	147	2.581
03	002	Aristoteles	1.263	166	2.092
17	022	Kant, Immanuel	847	130	2.652
17	115	"Aetas kantiana"	373	341	28
18	005	Hegel, Georg Wilhelm Friedrich	769	66	1.877
20b	0000	Lingua italiana (studi generali)	693	259	1.245
20e	0000	Estetica (studi generali)	595	205	1.518
20i	0001	Husserl, Edmund	626	64	2.206
20i	0002	Heidegger, Martin	466	95	1.673
20l	0000	Lingua inglese (studi generali)	627	310	657
20q	0605	Psicologia	484	339	739

Tabella 5

tali, i numeri di catena che appartengono a questo *cluster* sono responsabili di quasi il 18 per cento dei prestiti (17.268 su 102.885) e di oltre il 20 per cento delle nuove acquisizioni (2.816 su 13.705).

Una volta individuati i *cluster*, è possibile approfondire ulteriormente l'analisi (Tabella 5). Nel caso, ad esempio, di 17.115, una stella, il numero dei prestiti è molto basso se confrontato con quello degli altri, mentre quello delle nuove acquisizioni è più alto. "Aetas kantiana" è una collana di 271 volumi (corrispondenti a 373 tomi) pubblicati nel 2007 che la Biblioteca di Filosofia ha terminato di acquistare nel 2011. Questo spiega la peculiarità di questo numero di catena. Anche tra le mucche da latte ci sono numeri di catena le cui caratteristiche sono in parte differenti da quelle degli altri. È il caso, ad esempio, di 05 036 ("Sources chrétiennes"): pur comprendendo 229 volumi, non ha avuto nessuna nuova acquisizione nel periodo considerato e l'ultimo prestito risale al 2017. L'interesse verso la collana delle Editions du cerf dedicata agli scrittori cristiani dei primi secoli sembra quindi essersi esaurito, dato che tra il 2009 e il 2018 al suo

interno sono stati pubblicati oltre 80 nuovi volumi.

## Conclusioni

Le biblioteche accademiche hanno una lunga tradizione nella raccolta e nell'elaborazione dei dati. Nel 2000 è stato costituito il Gruppo interuniversitario per il monitoraggio dei sistemi bibliotecari di ateneo (GIM) che ha realizzato tre indagini nazionali.<sup>35</sup> Il progetto Good Practice, avviato nel 1999 all'interno del Dipartimento di ingegneria gestionale del Politecnico di Milano allo scopo di misurare l'efficienza delle amministrazioni universitarie, ha una parte specificamente dedicata ai sistemi bibliotecari di ateneo.<sup>36</sup> Esistono poi due standard ISO dedicati alle misure e agli indicatori delle biblioteche.<sup>37</sup> L'introduzione del ciclo della performance nelle università – avvenuta a seguito del Decreto legislativo 150

del 27 ottobre 2009, Attuazione della Legge 15/2009 in materia di ottimizzazione della produttività del lavoro pubblico e di efficienza e trasparenza delle pubbliche amministrazioni – ha in parte modificato la situazione per quanto riguarda la raccolta e l'elaborazione dei dati.

La performance amministrativa è la misura delle prestazioni di supporto a ogni attività che si realizza all'interno dell'università. L'esperienza internazionale ha dimostrato che l'introduzione di efficaci sistemi di performance nelle organizzazioni pubbliche complesse produce notevoli miglioramenti, tanto nelle prestazioni generali dell'ente, quanto nella qualità del lavoro delle persone coinvolte.<sup>38</sup>

Il ciclo della performance si articola in tre diversi momenti, che devono replicarsi di anno in anno in modo coerente e cumulativo: un atto di programmazione (il Piano integrato della performance); un'attività di monitoraggio ed eventuale correzione della programmazione; una valutazione dei risultati ottenuti, opportunamente rendicontata. Il Piano integrato della performance deve essere costruito seguendo due prin-

cipi di fondo: l'ancoraggio indissolubile della performance amministrativa con le missioni istituzionali dell'università e con le risorse finanziarie necessarie per perseguirle; l'assunzione della performance come concetto guida intorno al quale si collocano le diverse prospettive della trasparenza (concepita come obbligo di rendicontazione a studenti, famiglie, imprese e territorio) e della prevenzione alla corruzione (intesa come attività di contrasto e mitigazione dei rischi derivanti da comportamenti inappropriati e illegali).<sup>39</sup>

Le biblioteche, come tutte le altre componenti dell'amministrazione universitaria, sono quindi chiamate ad adottare un approccio che punti al miglioramento continuo dei processi che si svolgono al loro interno e consenta di misurare l'avanzamento dei nuovi progetti, comunicando i risultati a tutti gli interessati. Per farlo è necessario disporre di dati affidabili e continuamente aggiornati che possano guidare le decisioni.<sup>40</sup> Queste informazioni devono essere rappresentate attraverso grafici, che rendono possibile far comprendere anche ai portatori di interesse la distribuzione dei valori, la presenza di valori estremi e le interrelazioni all'interno dell'insieme dei dati.<sup>41</sup>

Tornando alla Biblioteca di Filosofia, l'esame dei cluster e delle caratteristiche utilizzate per raggruppare i numeri di catena permette di farsi un'idea esatta della situazione in cui si trova la collezione. I bibliotecari, in collaborazione con gli studenti e i docenti, hanno quindi tutti gli elementi per mettere a punto un nuovo schema di collocazione e per decidere quali settori possono essere sviluppati e quali possono essere dismessi, eventualmente collocando in magazzino i volumi.

L'analisi dei dati si è rivelata uno strumento in grado di mettere in luce punti di forza e debolezze della collezione e del catalogo fornendo gli strumenti per programmare interventi mirati e pensare azioni di ampio respiro che garantiscano una prospettiva di crescita organica e una fruizione efficace.

I bibliotecari sono stati dei *data scientist* già prima che venisse coniato il termine. Sin dalle origini, infatti, il loro lavoro è consistito nel raccogliere grandi quantità di dati, normalizzarli e presentarli in modo che potessero essere utilizzati al meglio. I progressi dell'intelligenza artificiale e della tecnologia rendono possibile svolgere queste attività con sempre maggior efficacia. È necessario quindi tornare alle origini, riappropriandosi delle collezioni e dei cataloghi.

Il vantaggio di utilizzare questi nuovi strumenti, il cui

impiego è sempre più alla portata anche dei non specialisti, avrebbe anche la conseguenza di riavvicinare chi studia o fa ricerca, dato che il bibliotecario sarebbe visto come un esperto nel trattamento dei dati bibliografici, dai quali è in grado di estrarre informazioni che non sono immediatamente accessibili a tutti.<sup>42</sup>

## NOTE

<sup>1</sup> MAURIZIO VIVARELLI, *La costruzione delle raccolte. Teorie e tecniche per lo sviluppo e la gestione delle collezioni*, in *Biblioteconomia: principi e questioni*, a cura di Giovanni Solimine e Paul Gabriele Weston, Roma, Carocci, 2007, p.3959, p. 41-43.

<sup>2</sup> SARA DINOTOLA, *La carta delle collezioni tra principi consolidati e nuovi scenari: qualche riflessione*, "Biblioteche oggi Trends", 4 (2018), 2, p. 6-17, p. 10-14.

<sup>3</sup> PEGGY JOHNSON, *Collection development and management*, Chicago, American Library Association, 2013<sup>3</sup>, p. 1.

<sup>4</sup> KAREN C. KOHN, *Collection evaluation in academic libraries. A practical guide for librarians*, Lanham, Rowman & Littlefield, 2015, p. 2-4.

<sup>5</sup> JIM COLEMAN, *The RLG Conspectus*, "The Acquisitions Librarian", 4 (1992), 7, p. 25-43.

<sup>6</sup> SARA DINOTOLA, *La carta delle collezioni tra principi consolidati e nuovi scenari: qualche riflessione*, cit., p. 16.

<sup>7</sup> ANDREA GIARDINA, *Esplosione di tardo antico*, "Studi storici", 40 (1999), 1, p. 157-180.

<sup>8</sup> L'indirizzo del sito è [sidoapol.org](http://sidoapol.org).

<sup>9</sup> ROB JOHNSON, ANTHONY WATKINSON, MICHAEL MABE, *The STM Report an overview of scientific and scholarly publishing*, The Hague, International Association of Scientific, Technical and Medical Publishers, 2018, p. 5

<sup>10</sup> ESTHER LANDHUIS, *Information overload. How to manage the research-paper deluge? Blogs, colleagues and social media can all help*, "Nature", 535 (2016), 21 luglio 2016, p. 457-458.

<sup>11</sup> BRIAN GODSEY, *Think like a data scientist. Tackle the data science process step-by-step*, Shelter Island, Manning, 2017.

<sup>12</sup> MANUELA D'URSO, *Stack management: la gestione delle collezioni sugli scaffali in biblioteca*, "JLIS.it.", 2 (2011), 1.

<sup>13</sup> Tutte le informazioni su WorldShare Collection Evaluation sono disponibili all'indirizzo [www.oclc.org/en/collection-evaluation.html](http://www.oclc.org/en/collection-evaluation.html).

<sup>14</sup> BRUCE MASSIS, *Data-driven decision-making in the library*, "New library world", 117 (2016), 1/2, p. 131-134.

<sup>15</sup> Ringrazio Laura Frigerio, responsabile della Biblioteca di Filosofia, e Yuri Gallo, che ci lavora, per avermi descritto la situazione in cui si trova attualmente la loro collezione

e avermi messo a disposizione i dati sui prestiti e le nove acquisizioni.

<sup>16</sup> LILIANA GRASSI, *La Ca' Granda. Storia e restauro*, Milano, Università degli studi di Milano, 1958.

<sup>17</sup> GIULIANA SAPORI, *Note sulla Biblioteca delle facoltà di Giurisprudenza, Lettere e Filosofia dell'Università degli Studi di Milano* (il testo è consultabile sul sito dell'Università all'indirizzo [www.sba.unimi.it/files/bglf/noteBGLF.pdf](http://www.sba.unimi.it/files/bglf/noteBGLF.pdf)).

<sup>18</sup> I libri pubblicati prima del 1945 sono esclusi dal prestito anche se inseriti in una delle collocazioni della serie generale.

<sup>19</sup> LAURA FRIGERIO, *From disorder to order: a challenge for the philosopher and the librarian*, "Knowledge organization", 2009, 2/3, p. 150-159.

<sup>20</sup> "B.2.1.1 Collection Turnover: The total number of loans in the specified collection during a specified period of time, normally one year, divided by the total number of documents in the collection." (ISO 11620:2014: *Information and documentation - Library performance indicator*, Genève, ISO, 2014).

<sup>21</sup> ROSWITA POLL, PETER TE BOEKHORST, *Measuring quality. Performance measurement in libraries*, 20072, p. 131.

<sup>22</sup> *Ibidem*.

<sup>23</sup> ISO 2789:2013: *Information and documentation - International library statistics*, Genève, ISO, 2014.

<sup>24</sup> PAOLO CASALEGNO, *Filosofia del linguaggio: un'introduzione*, Roma, Carocci, 1997.

<sup>25</sup> Per chi non avesse familiarità con questo tipo di scala ricordiamo che  $10^0$  vale 1,  $10^1$  10,  $10^2$  100,  $10^3$  1.000 e  $10^4$  10.000.

<sup>26</sup> DAVID WOLPERT, *The lack of a priori distinctions between learning algorithms*, "Neural computation", 1996, p. 13411390.

<sup>27</sup> IAN GOODFELLOW, YOSHUA BENGIO, AARON COURVILLE, *Deep learning*, Cambridge, Massachusetts, The MIT Press, 2016.

<sup>28</sup> La frase è stata pronunciata da Yann LeCun nel corso del suo intervento alla trentesima Conference on Neural Information Processing Systems svoltasi a Barcellona tra il 5 e il 10 dicembre 2016.

<sup>29</sup> *Unsupervised learning and clustering*, in RICHARD O. DUDA, PETER E. HART, DAVID G. STORK, *Pattern classification*, New York, Wiley, 20012, p. 517-599.

<sup>30</sup> STUART P. LLOYD, *Least squares quantization in PCM*, "IEEE transactions on information theory", 28 (1982), 2, p. 129137 (la prima versione dell'articolo risale al 1957).

<sup>31</sup> EDWARD W. FORGY, *Cluster analysis of multivariate data: efficiency versus interpretability of classifications*, "Biometrics", 21 (1965), 3, p. 768-769.

<sup>32</sup> SHAI SHALEV-SHWARTZ, SHAI BEN-DAVID, *Understanding ma-*

*chine learning. From theory to algorithms*, New York, Cambridge University Press, 2014, p. 264-277.

<sup>33</sup> AURÉLIEN GÉRON, *Handson machine learning with Scikit-Learn, Keras, and TensorFlow. Concepts, tools, and techniques to build intelligent systems*, Sebastopol, O'Reilly, 2019<sup>2</sup>.

<sup>34</sup> BRUCE HENDERSON, *The product portfolio*, "BCG Henderson Institute", 1<sup>st</sup> January 1970, disponibile sul sito del Boston Consulting Group all'indirizzo [www.bcg.com](http://www.bcg.com). In questa prima versione i cani sono chiamati animali domestici (*pet*).

<sup>35</sup> Il sito di GIM è raggiungibile all'indirizzo [www.gimsba.it](http://www.gimsba.it).

<sup>36</sup> Le informazioni sul progetto sono reperibili all'interno di una sezione del sito della School of Management del Politecnico di Milano all'indirizzo [www.som.polimi.it/en/research/research\\_lines/good\\_practice](http://www.som.polimi.it/en/research/research_lines/good_practice).

<sup>37</sup> INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, *ISO 2789:2013 Information and documentation - International library statistics*, ISO, Genève, 2013; International Organization for Standardization, *ISO 11620:2014 Information and documentation - Library performance indicators*, ISO, Genève, 2014.

<sup>38</sup> AGENZIA NAZIONALE DI VALUTAZIONE DEL SISTEMA UNIVERSITARIO E DELLA RICERCA, *Linee guida per la gestione integrata del ciclo della performance delle università statali italiane*, luglio 2015, p. 4 (il documento è disponibili sul sito dell'ANVUR all'indirizzo [anvur.it](http://anvur.it)).

<sup>39</sup> *Ivi*, p. 8-11, passim.

<sup>40</sup> LESLEY S. J. FARMER, ALAN M. SAFER, *Library Improvement Through Data Analytics*, Chicago, ALA Neal-Schuman, 2016.

<sup>41</sup> VJIAI KOTU, BALA DESHPANDE, *Data Science. Concepts and Practice*, Burlington, Morgan Kaufmann, 2019, p. 25.

<sup>42</sup> MATT BURTON, LIZ LYON, CHRIS ERDMANN, BONNIE TIJERINA, *Shifting to data savvy: the future of data science in libraries*, Institute for Museum and Library Services, 2017.

## ABSTRACT

The article examines the library collection of the Philosophy Library of the University of Milan. For this purpose, methodologies used in unsupervised learning have been used. The result was the grouping of the different sections of the book collection into five clusters analyzed through the matrix developed by the Boston Consulting Group. The results of the survey provided the basis for launching a series of actions in order to increase the use of the collection.

DOI: 10.3302/0392-8586-202001-016-1