

«Dietro è la casa, davanti a noi il mondo»: da UNIMARC a MARC₂₁ all'Università di Milano-Bicocca

di Alessandra Moi, Silvana Re, Francesca Verga

L'Università degli studi di Milano-Bicocca

L'Università di Milano-Bicocca, fondata nel 1998, in trent'anni ha raggiunto dimensioni considerevoli: a fine 2021 ospitava 39.011 studenti e dottorandi, 1.269 tra professori e ricercatori, 849 tecnici e amministrativi, 416 assegnisti di ricerca.

La Biblioteca di ateneo è stata concepita fin dall'inizio come una singola istituzione, articolata fisicamente in tre sedi distinte (Medicina, Centrale e Scienze).

La sede Centrale della biblioteca ospita inoltre il Polo di archivio storico (PAST)¹, che conserva i fondi archivistici di psicologi e psichiatri raccolti dall'ASPI - Archivio storico della psicologia italiana² e tutti gli archivi storici acquisiti dall'Ateneo grazie a lasciti e donazioni.

I bibliotecari sono 37 ed è comune per una stessa persona occuparsi di più mansioni, a volte anche in sedi diverse: questo contribuisce a rafforzare la percezione della biblioteca come un organismo unico.

Tutte le sedi usano gli stessi software e le procedure di acquisizione, catalogazione e circolazione sono condivise, indipendentemente dalla localizzazione e dalla disciplina di appartenenza del materiale.

A fine 2020, il patrimonio della biblioteca era suddiviso come segue:
- 294.183 monografie a stampa (tra cui 1.345 libri antichi);

ALESSANDRA MOI, Università degli studi di Milano-Bicocca, e-mail alessandra.moi@unimib.it.

SILVANA RE, Università degli studi di Milano-Bicocca, e-mail silvana.re@unimib.it.

FRANCESCA VERGA, Università degli studi di Milano-Bicocca, e-mail francesca.verga@unimib.it.

Il titolo dell'articolo prende spunto dal passo «Home is behind, the world ahead» tratto da *The fellowship of the ring* di J.R.R. Tolkien (1954). Si ringraziano il direttore Maurizio di Girolamo e tutti i colleghi della Biblioteca di ateneo dell'Università di Milano-Bicocca per lo straordinario lavoro fatto.

Ultima consultazione siti web: 2 settembre 2022.

1 Il PAST - Polo di archivio storico dell'Università degli studi di Milano-Bicocca (<<https://www.past.biblio.unimib.it/>>) conserva tutti gli archivi storici acquisiti dall'Ateneo grazie a lasciti e donazioni e i fondi archivistici di psicologi e psichiatri raccolti dall'ASPI - Archivio storico della psicologia italiana.

2 Il centro ASPI - Archivio storico della psicologia italiana (<<https://www.aspi.unimib.it/>>) dell'Università degli studi di Milano-Bicocca individua, raccoglie e mette a disposizione online gli archivi storici degli scienziati della mente (psicologi, psichiatri, neurologi ecc.) attivi in Italia nell'Ottocento e nel Novecento.



- 57.590 annate di periodici a stampa;
- 119 banche dati;
- 292.987 e-book;
- 74.800 periodici elettronici;
- 2.371 tra CD-ROM, DVD e altro materiale audiovisivo;
- 309 carte geografiche e altri materiali.

Caratteristiche dei sistemi gestionali precedenti

Fino alla gara (avviata il 9 giugno 2020), il posseduto della biblioteca è stato gestito con tre software diversi, oltre a strumenti di *office automation* come Microsoft Excel.

Il software Aleph di Ex Libris³ gestiva l'acquisizione e la catalogazione del materiale a stampa, la circolazione e i servizi interbibliotecari. La catalogazione era basata sul formato UNIMARC. Per una scelta consapevole, il materiale elettronico non è mai stato catalogato in Aleph.

Il link resolver SFX⁴, sempre di Ex Libris, gestiva e-journal ed e-book dal punto di vista della copertura e dell'accesso, ma non aveva funzioni di acquisizione o catalogazione.

Gli acquisti di risorse elettroniche, comprese le banche dati, venivano quindi registrati in una serie di fogli di calcolo, mentre i record catalografici venivano scaricati dal sito dell'editore o da altre fonti, salvati in locale e convertiti in MARCXML basato su MARC21.

La visibilità dell'intero posseduto era garantita dal discovery tool Curiosone, basato su EBSCO Discovery Service (EDS)⁵.

EDS conteneva sia record forniti da banche dati partner di EBSCO, relativi soprattutto ad articoli, sia record provenienti da caricamenti del nostro posseduto.

Questi ultimi erano automatici e giornalieri per quanto riguardava i record UNIMARC, mentre il materiale MARCXML veniva caricato manualmente dopo ogni acquisto oppure a scadenze fisse (in genere mensili), a seconda della tipologia di acquisizione (acquisto singolo o abbonamento).

In sede di configurazione si era deciso di mostrare agli utenti solo i record bibliografici effettivamente posseduti, usando un filtro basato su altri caricamenti automatici da Aleph e da SFX.

In aggiunta al materiale della biblioteca, i record provenienti dagli archivi PAST e ASPI e dal *repository* di Ateneo (Bicocca open archive, BOA⁶) venivano caricati automaticamente tramite il protocollo OAI-PMH⁷.

3 <<https://exlibrisgroup.com/it>>.

4 SFX è stato il primo link resolver basato sul protocollo OpenURL, sviluppato alla fine degli anni Novanta in Belgio dalla Ghent University (Wendy C. Robertson; Paul A. Soderdahl, *Everything you always wanted to know about SFX but were afraid to ask*, «The serials librarian», 47 (2004), n. 1-2, p. 129-138, DOI: 10.1300/J123V47N01_09; Herbert Van de Sompel; Oren Beit Arie, *Open linking in the scholarly information environment using the OpenURL Framework*, «New review of information networking», 7 (2001), n. 1, p. 59-76, DOI: 10.1080/13614570109516969).

5 <<https://www.ebsco.com/>>.

6 IRIS è il sistema di gestione della ricerca dell'Università degli studi di Milano-Bicocca. Il *repository* BOA (<<https://boa.unimib.it/>>) è il modulo del sistema dedicato alla raccolta e alla disseminazione della produzione scientifica di ateneo.

7 Il protocollo OAI-PMH (Open access initiative protocol for metadata harvesting, <<https://www.openarchives.org/pmh/>>) permette di esporre i metadati nativi di un *repository* secondo un set condiviso di attributi, in modo che altri sistemi possano interrogarli e importarli (Marshall Breeding, *Understanding the protocol for*

Nonostante l'assistenza fornita dai produttori dei software, la gestione di questi processi risultava molto onerosa: il posseduto digitale doveva essere tenuto aggiornato in più punti, i caricamenti automatici dovevano comunque essere monitorati e l'interfaccia del discovery tool richiedeva a sua volta aggiornamenti e manutenzione.

Infine, le statistiche delle varie risorse e servizi (in particolare le statistiche COUNTER) dovevano essere scaricate singolarmente e gestite con fogli di calcolo.

La scelta di passare a una *library services platform* (LSP)⁸ è quindi imputabile essenzialmente a due motivi:

- la mancanza di un ERMS⁹ per l'amministrazione, l'acquisizione e la catalogazione delle risorse elettroniche, a fronte di un grande incremento di queste ultime, anche precedente alla pandemia;

- la crescente difficoltà di gestire un posseduto eterogeneo per software di provenienza e formato, facendolo confluire in un unico discovery tool.

Si è deciso quindi di cercare un software che permettesse di operare sempre in un unico ambiente, riutilizzando gli stessi dati catalografici, e che offrisse per la prima volta un ambiente di gestione delle risorse digitali integrato con quello delle risorse a stampa.

È noto che alcune di queste piattaforme non gestiscono il formato UNIMARC, ma solo il formato MARC21: si è ritenuto però che la difficoltà temporanea del cambio di formato fosse compensata dal fatto che MARC21 è il formato più diffuso a livello globale e viene aggiornato e mantenuto in maniera cooperativa da alcune tra le più importanti biblioteche di area angloamericana¹⁰.

A dicembre 2020 la gara è stata aggiudicata a OCLC¹¹, che supporta solo MARC21. La biblioteca si è preparata quindi al cambio di formato¹², tappa fondamentale della

metadata harvesting of the open archives initiative, «Computers in libraries», 22 (2002), n. 8, p. 24-29; Herbert Van de Sompel; Carl Lagoze, *The Santa Fe convention of the Open archives initiative*, «D-Lib magazine», 6 (2000), n. 2, <<https://dspace.library.uu.nl/bitstream/handle/1874/3142/VandeSompelD-Lib2000SantaFe.htm?sequence=2&isAllowed=y>>).

8 Una *library services platform* (o *library management platform*) si differenzia da un ILS (*integrated library system*) per l'integrazione completa tra i flussi di lavoro delle risorse elettroniche e quelli delle risorse a stampa. È inoltre *cloud-based*, mentre gli ILS avevano un'architettura client-server (Sara Dinotola, *I sistemi per la gestione delle risorse elettroniche (seconda parte): le library service platforms (LSP)*, «AIB studi», 56 (2016), n. 2, p. 205-218, DOI: 10.2426/aibstudi-11412).

9 Un *electronic resource management system* (ERMS) è un sistema gestionale per il ciclo di vita delle risorse elettroniche. Per approfondimenti sugli ERMS si rimanda a: Elsa K Anderson, *Electronic resource management systems: a workflow approach*. Chicago: ALA TechSource, 2015; Anne C. Elguindi; Kari Schmidt, *Electronic resource management: practical perspectives in a new technical services model*. Oxford: Chandos, 2012.

10 Si tratta nello specifico della Library of Congress (Network Development and MARC Standards Office), della Library and Archives Canada (Content Management Division) e della British Library (Bibliographic Standards and Systems). Per ulteriori approfondimenti si rimanda al sito ufficiale del formato MARC21: <<https://www.loc.gov/marc/bibliographic/bdintro.html>>.

11 OCLC (<<https://www.oclc.org/en/home.html>>) è un'organizzazione cooperativa non profit di biblioteche, con sede negli Stati Uniti a Dublin, Ohio. Il prodotto più noto di OCLC è WorldCat, un catalogo bibliografico collettivo, che raccoglie il posseduto di 10.000 biblioteche in tutto il mondo.

12 Analoghe esperienze sul cambio di formato sono state, soprattutto nell'area europea, abbastanza significative e molto ben documentate. Per eventuali confronti con la situazione affrontata dall'Università di Milano-Bicocca si rimanda a: Gordana Rudi; Duan Surla, *Conversion of bibliographic records to MARC 21 format*, «The

migrazione, che è iniziata con l'aggiudicazione efficace il 15 febbraio 2021 e si è conclusa a fine ottobre 2021, con la messa online del nuovo discovery.

I formati MARC

Evoluzione storica e caratteristiche

Il primo tentativo di realizzare un formato *machine-readable* finalizzato allo scambio dei dati bibliografici va ricondotto al MARC pilot project, intrapreso dalla Library of Congress nel 1966 con la collaborazione di 16 istituti bibliotecari partecipanti, che si concluse con l'elaborazione del MARC II format nel 1968¹³. La rapida diffusione del MARC come formato principale per lo scambio dei record, bibliografici e non, è da imputare alla sua rigorosa e complessa strutturazione, la quale permette di esprimere una grande quantità e varietà di informazioni attraverso l'utilizzo dei cosiddetti 'campi', funzionali alla codifica delle informazioni da parte delle macchine. Nonostante le indubbie differenze, i vari formati MARC rispondono a una medesima strutturazione, derivante dallo standard internazionale ISO 2709 e dal corrispettivo American national standard for information interchange ANSI/NISO Z39.2.

Nel formato MARC le informazioni sono organizzate in una successione di campi distinti nelle due diverse tipologie a lunghezza fissa e variabile, a loro volta ulteriormente articolati in una successione di etichette (*tag*), indicatori (*indicator*) e sottocampi (*subfield*).

Il MARC ha goduto e gode tuttora di larga fortuna essendo divenuto il formato in assoluto più utilizzato nella trasmissione delle informazioni bibliografiche. La presenza di centinaia di campi, dal 000 al 999, consente di strutturare informazioni complesse, di frammentarle e renderle elaborabili dalla macchina, a prescindere dalla loro formulazione in codici e standard catalografici specifici¹⁴. L'interoperabilità del dato ha garantito un costante riuso, con notevole profitto sia in termini di qualità dei dati stessi (acquisizione di dati provenienti da produttori autorevoli) sia di risparmio economico, favorendo la catalogazione derivata.

*UNIMARC e MARC21*¹⁵

Pur partendo da una comune conformazione, i due formati UNIMARC e MARC21 sono contraddistinti da significative differenze che investono vari livelli: storico, gestionale e di strutturazione delle informazioni.

electronic library», 27 (2009), n. 6, p. 950-967, DOI: 10.1108/02640470911004057; R.W. Hill, *Changing the record: a concise guide to the differences between the UKMARC and MARC 21 bibliographic formats*. Boston: The British Library, 2002.

13 Per approfondimenti sulla storia del formato MARC: Henriette D. Avram, *The MARC pilot project: final report*. Washington: Library of Congress, 1968; Michele Seikel; Thomas Steele, *How MARC has changed: the history of the format and its forthcoming relationship to RDA*, «Technical services quarterly», 28 (2011), n. 3, p. 322-334, DOI: 10.1080/07317131.2011.574519; Karen M. Spicher, *The development of the MARC format*, «Cataloging & classification quarterly», 21 (1996), n. 3-4, p. 75-90, DOI: 10.1300/J104v21n03_06.

14 Sally McCallum, *MARC: keystone for library automation*, «IEEE annals of the history of computing», 24 (2002), n. 2, p. 34-49, DOI: 10.1109/MAHC.2002.1010068.

15 Le immagini contenute nel presente paragrafo sono tratte da SBN, per gli esempi sul formato UNIMARC, e dal discovery tool Prometeo dell'Università di Milano-Bicocca, per gli esempi sul formato MARC21.

Per quanto riguarda la loro genesi storica, l'UNIMARC e il MARC21 non solo sono nati in periodi abbastanza distanti, ma altresì in virtù di esigenze significativamente diverse.

L'UNIMARC¹⁶ (Universal MARC format) venne pubblicato nel 1977 dall'IFLA con l'obiettivo di ricomprendere in sé i vari formati MARC a livello internazionale. L'UNIMARC è stato adottato in numerosi paesi europei, tra i quali l'Italia, ed è funzionale al trattamento dei record bibliografici e di *authority*, delle *holding* e dei dati di classificazione¹⁷.

La 'nascita' del MARC21¹⁸, invece, scaturì da una collaborazione tra la Library of Congress e la National Library of Canada, finalizzata all'armonizzazione tra i formati USMARC e CANMARC. Nel 1999 vennero pubblicati tutte e cinque i formati specifici che ancora oggi compongono il MARC21: Format for authority data; Format for bibliographic data; Format for classification data; Format for community information; Format for holdings data¹⁹.

UNIMARC e MARC21, per la loro diversa origine e il diverso ambito d'uso, sono quindi fortemente vincolati ai rispettivi approcci catalografici: europeo per l'UNIMARC, angloamericano per il MARC21.

Per questo motivo spesso si fatica a trovare un possibile allineamento sia nel significato, sia nella composizione dei campi e delle informazioni contenute. Tra le principali differenze si trovano²⁰:

- strutturazione dei campi per le informazioni codificate: le informazioni codificate sono gestite, sia nell'UNIMARC sia nel MARC21, attraverso campi a lunghezza fissa, riconducibili rispettivamente al blocco 1XX e 00X. Tuttavia, ciò che nell'UNIMARC viene espresso tramite l'utilizzo di più campi, alcuni dei quali specifici per ciascuna tipologia di materiale, nel MARC21 deve essere ricondotto ai soli campi 006, 007 e 008. Dei tre, il campo 008 è il più complesso, in quanto i valori esprimibili variano a seconda della risorsa catalogata;

- gestione delle responsabilità primarie: riprendendo un'impostazione della scheda cartacea, nel formato MARC21 si distingue nettamente tra responsabilità principale e responsabilità secondarie. L'autore principale, infatti, è espresso all'interno del blocco 1XX (campi 100/110/111 per gli accessi alle responsabilità di persone, enti e congressi) e precede sempre i campi della descrizione bibliografica; viceversa, le responsabilità secondarie sono espresse nel blocco 7XX. Tale distinzione è totalmente assente nel formato UNIMARC, nel quale sia le responsabilità principali sia quelle secondarie sono inserite nel blocco 7XX;

16 <<https://www.ifla.org/publications/unimarc-formats-and-related-documentation>>.

17 Hugo Manguinhas; Nuno Freire; José Borbinha, *Management of metadata standards: the case of UNIMARC*. In: *Digital libraries: achievements, challenges and opportunities: 9th International conference on Asian digital libraries, ICADL 2006 Kyoto, Japan, November 27-30*, edited by Shigeo Sugimoto [et al.]. Berlin: Springer, 2006, p. 264-273, DOI: 10.1007/11931584_29.

18 <<http://www.loc.gov/marc/annmarc21.html>>.

19 <<https://www.loc.gov/marc/>>.

20 Per ulteriori differenze tra i due formati si veda: Zeno Tajoli, *Da Unimarc a MARC21*, «Bollettino del CILEA», 2004, n. 95, p. 27-28, <<http://bollettino.cilea.it/article/view/6724>>.

- 100 \$a20140530d2014 |||0|tac50ba
- 101 \$aita
- 102 \$aIT
- 181 1\$&z01\$ai \$bbooxe
- 182 1\$&z01\$an
- 183 1\$&z01\$anc\$2RD\$acARRIER
- 200 1 \$aStoria dell'antropologia cinese\$fHu Hongbao, Wang, Jianmin, Zhang Haiyang&gedizione italiana a cura di Andrea E. Pia>raduzione di Andrea E. Pia ... [et al.]
- 210 \$aFirenze&\$SEID\$d2014
- 215 \$aXV, 296 p.\$d26 cm
- 300 \$aNome del curatore in copertina: Andrea Pia
- 410 0\$1001ITNCCUUFN0452349\$12001 \$aCollana di antropologia
- 606 \$aAntropologia&xCina\$2FI \$3ITNCCUL01C298917
- 676 \$a301.0951&cSOCIOLOGIA E ANTROPOLOGIA. Cina\$v23
- 700 0\$aHu Hongbao&3ITNCCUFCFV317173
- 701 0\$aWang, Jianmin&3ITNCCUFCFV317181
- 701 0\$aZhang Haiyang&3ITNCCUFCFV317183
- 702 1\$aPia\$b, Andrea E.\$3ITNCCUFCFV317171

Figura 1 – Tag 700/701 nell'UNIMARC

100 1# \$a Hu, Hongbao \$4 aut
245 00 \$a Storia dell'antropologia cinese / Sc Hu Hongbao, Wang Jianmin e Zhang Haiyang ; edizione italiana a cura di Andrea E. Pia ; traduzione di Andrea E. Pia ... [et al.].
260 ## \$a Firenze : Sb SEID, \$c 2014.
300 ## \$a xv, 296 p. ; \$c 26 cm.
336 ## \$a testo \$b txt \$2 rdacontent/ita
337 ## \$a senza mediazione \$b n \$2 rdamedia/ita
490 1# \$a Collana di antropologia
504 ## \$a Con bibliografia (p. 267-290) e indice.
650 04 \$a Antropologia \$z Cina \$x Storia.
700 1# \$a Wang, Jianmin \$4 aut
700 1# \$a Zhang, Haiyang \$4 aut
700 1# \$a Pia, Andrea E \$4 tr1
830 #0 \$a Collana di antropologia (Seid)

Figura 2 – Tag 100/700 nel MARC21

- formulazione nome/titolo: esiste una parziale differenza tra i due formati relativi alla formulazione dei titoli delle opere (o titoli uniformi). Nel MARC21, i tag primari 130 e 240 per l'indicazione dei titoli delle opere non sono ripetibili, quindi è necessario l'utilizzo di appositi tag appartenenti al blocco 7XX. Mentre per le opere prive di responsabilità la questione si risolve attraverso l'uso di uno o più campi 730, per le opere con responsabilità principale il MARC21 prevede una vera e propria combinazione tra nome dell'autore e titolo dell'opera all'interno dello stesso campo 700/710/711. Nell'UNIMARC, invece, più titoli di opere all'interno dello stesso record possono essere espressi o tramite l'uso di più tag 500 (campo ripetibile) oppure tramite il tag 576: quest'ultimo nello specifico condivide con la struttura nome/titolo del MARC21 la combinazione tra nome dell'autore e titolo dell'opera, seppur con alcune differenze.

- 200 1 \$aDonne contro\$fa cura di Ettore Barelli
- 210 \$aMilano&\$cEdizioni scolastiche B. Mondadori&\$dc1998
- 215 \$a125 p.\$d24 cm
- 300 \$aContiene: Antigone / Sofocle . Casa di bambola / Henrik Ibsen
- 410 0\$1001ITNCCUFCFV0532614\$12001 \$aSpecchi teatro\$v15
- 423 0\$1001ITNCCUURAV0162015\$12001 \$aCasa di bambola\$fHenrik Ibsen. -\$1700 1\$aIbsen\$b, Henrik&3ITNCCUFCFV008640\$4070
- 500 11\$aAntigone&3ITNCCUBVEE007856\$9Sophocles
- 500 11\$a'Et %dukkkehjem&3ITNCCUMIL041094\$9Ibsen, Henrik
- 702 1\$aBarelli\$b, Ettore&3ITNCCUFCFV024269

Figura 3 – Doppio tag 500 nell'UNIMARC

245 10 \$a Donne contro ; \$b Antigone / \$c Sofocle ; Casa di bambola / Henrik Ibsen ; a cura di Ettore Barelli.
260 ## \$a Milano : \$b Edizioni scolastiche B. Mondadori, \$c ©1998.
300 ## \$a 125 p. ; \$c 24 cm.
336 ## \$a testo \$b txt \$2 rdacontent/ita
337 ## \$a senza mediazione \$b n \$2 rdamedia/ita
490 0# \$a Specchi teatro
505 00 \$t Antigone / \$r Sofocle
505 80 \$t Casa di bambola / \$r Henrik Ibsen
653 0# \$a Letteratura drammatica
700 1# \$a Barelli, Ettore. \$4 edt
700 02 \$a Sophocles. \$t Antigone. \$l Italiano
700 12 \$a Ibsen, Henrik. \$t Dukkehjem. \$l Italiano

Figura 4 – Combinazione nome-titolo nel MARC21

Tali distinzioni, che sono assolutamente parziali rispetto al più ampio complesso di incongruenze tra i due formati, pongono delle sfide notevoli per chiunque si accinga ad attività di analisi, mappatura e migrazione tra i due formati.

Dall'UNIMARC al MARC21

Il contesto della migrazione

Il passaggio verso un nuovo sistema gestionale richiede sia impegno ed energie sia consapevolezza delle differenze rispetto al contesto precedente. Nella transizione dal pregresso sistema Aleph a WorldShare Management Services (WMS) è stato necessario affrontare alcuni fondamentali cambiamenti, che non hanno investito il solo cambio di formato, bensì lo stesso approccio alle procedure catalografiche.

In primo luogo, è stato necessario comprendere come vengono gestiti i record bibliografici nel nuovo sistema: in WMS, infatti, è prevista la catalogazione per singola manifestazione dell'opera, in riferimento alla specifica lingua di catalogazione di ciascuna istituzione. In una situazione ideale, esisterebbe un record catalografico per singola manifestazione e per specifica lingua di catalogazione; naturalmente non è così, poiché le scelte catalografiche delle varie istituzioni che operano in una medesima lingua di catalogazione possono variare e dare luogo alla presenza di più record descrittivi per la stessa manifestazione.

Altro fondamentale aspetto di novità ha riguardato la struttura stessa del record bibliografico, diviso in WMS in 2 sezioni:

- il segmento BIB, detto anche *master record*, è condiviso con le altre biblioteche aderenti a WorldCat e contiene i principali campi descrittivi;
- il segmento LBD (*local bibliographic data*) contiene informazioni di specifica pertinenza della singola istituzione.

Al record bibliografico si aggiunge poi la sezione LHR (*local holdings record*), che contiene i dati della copia.

Mentre i dati contenuti nel *master record* possono essere modificati da qualunque partecipante a WMS, i record LBD e LHR sono accessibili alla sola istituzione che li ha creati.

In aggiunta a questi due aspetti, la biblioteca ha dovuto affrontare il 'salto' da un catalogo locale, totalmente gestito e gestibile in proprio, a un catalogo condiviso a livello globale, con la conseguenza di non poter gestire in piena autonomia i record bibliografici.

Ulteriore elemento di complessità è stato, infine, il cambio di formato da UNIMARC a MARC21, che ha richiesto un approfondito studio e un'ampia analisi dei metadati sia bibliografici sia di copia.

La bonifica dei dati

A dicembre 2018, ancora prima di pensare concretamente alla migrazione, la biblioteca aveva avviato una bonifica sui dati UNIMARC. La bonifica è stata svolta ancora in ambiente Aleph, parallelamente alle attività ordinarie, e si è conclusa a marzo 2021. Il suo scopo era non solo di facilitare una futura, eventuale migrazione, ma anche di garantire una certa coerenza interna al catalogo. La biblioteca infatti ha lavorato in Aleph/UNIMARC per circa vent'anni, in seguito a una migrazione da SBN avvenuta nel 2002: questo ha comportato una stratificazione delle scelte catalografiche. L'impostazione ereditata da SBN ha determinato i due interventi più significativi: la semplificazione delle monografie a livelli e la bonifica sui legami dei periodici.

Il lavoro sulle monografie a livelli è stato certamente il più complesso e ha richiesto diverse fasi di intervento. In primo luogo è stato necessario agire su tutte le catalogazioni a 'tre livelli' (record della monografia superiore, livello intermedio e livello inferiore), rimuovendo – tramite azioni in parte automatiche e in parte manuali – tutti i livelli intermedi e uniformando la catalogazione a soli due livelli. Inoltre sono stati eliminati tutti i record bibliografici con titolo non significativo (i cosiddetti "W" di SBN), preferendo l'utilizzo della nota di contenuto. Questo intervento è stato fatto anche nella consapevolezza che numerosi discovery tool e LSP non gestiscono correttamente le catalogazioni su più livelli. La scelta è stata così orientata al mantenimento della catalogazione a livelli solo in casi ben definiti, con una significativa revisione anche delle procedure catalografiche adottate.

Per quanto riguarda la bonifica dei legami sui periodici, l'attività ha interessato in particolar modo tutti quei legami orizzontali di successione o supplemento (esempio campi UNIMARC 430/440 e 421/422) importati dai reticoli di SBN, ma che all'interno di Aleph risultavano privi di significato in quanto non corrispondenti al possesso della biblioteca. Si è dunque provveduto alla rimozione manuale di questi campi nei record individuati grazie ad apposite estrazioni.

Oltre a questi due macrointerventi si sono poi aggiunte bonifiche più circoscritte, tra le quali la cancellazione, spesso automatizzata, di numerosi campi UNIMARC obsoleti o usati in maniera scorretta (esempio il campo UNIMARC 035 con l'indicazione del BID oppure il campo 791 per le forme varianti degli autori, entrambi importati da SBN). Interessante è stata invece la correzione, totalmente manuale, dei record bibliografici associati a un fondo personale acquisito dalla biblioteca. I record di questo fondo erano stati precedentemente importati tramite una procedura *batch* a partire da un minimale foglio di calcolo: per questo motivo erano privi di alcuni campi fondamentali, tra cui il Leader, e risultavano dunque invisibili agli utenti in quanto non gestiti dal discovery.

Tutte le attività di bonifica, dalla più complessa e faticosa fino alla più banale, hanno sicuramente agevolato il passaggio al nuovo formato: la partenza da una base dati relativamente coerente e omogenea si è, di fatto, rivelata fondamentale.

I passaggi di lavorazione

La definizione dei passaggi funzionali alla migrazione è stata portata avanti in stretta collaborazione con lo staff di OCLC e ha coinvolto il personale della biblioteca dalla stipula del contratto (aprile 2021) fino alla vera e propria entrata in produzione (ottobre 2021).

Il lavoro è stato articolato in una fase di pre-elaborazione e in una fase di caricamento dei dati.

La migrazione/importazione dei record bibliografici dal sistema Aleph/UNIMARC al sistema WMS/MARC21 è avvenuta attraverso il servizio di sincronizzazione, principalmente con le seguenti operazioni:

- l'aggiunta di nuovi record originali in WMS;
- l'individuazione di una possibile corrispondenza tra i record provenienti dal catalogo locale e i record già presenti.

In base a queste funzioni la migrazione/importazione prevedeva come risultato due possibili scenari: l'aggiunta dei record bibliografici in WMS, se le notizie bibliografiche non erano presenti nel database (*record added*), oppure lo schiacciamento dei record importati su notizie bibliografiche già esistenti (*record merged*).

In generale, la mappatura dei campi del record bibliografico per cui era possibile fare un'equivalenza automatica (ad esempio il campo 200 UNIMARC che corrisponde inequivocabilmente al campo 245 MARC21) è stata eseguita interamente da OCLC. La biblioteca si è invece concentrata sulla mappatura di quei metadati che potevano essere gestiti in modo personalizzato, ovvero gli LBD, e sui record master di tipo *added*. A partire da queste importanti premesse si è proceduto alla lavorazione dei record.

La fase di pre-elaborazione ha visto due momenti ben distinti: una parte teorica sulla conversione del formato e l'individuazione delle possibili problematiche nell'importazione (aprile-luglio 2021) e una parte operativa di test di caricamento (settembre-ottobre 2021). Durante il momento di analisi teorica sono state prese numerose decisioni, alcune strategiche ai fini dell'importazione e altre più di carattere operativo.

Scelta di grande importanza è stata quella di non migrare in WMS i record che in Aleph non erano visibili agli utenti, ad esempio i record soppressi (disinventariati, smarriti ecc.) oppure i record cosiddetti *on the fly*, creati per le catalogazioni temporanee. Anche sui record di copia sono state prese delle decisioni analoghe, con l'esclusione dalla migrazione di copie non valide (disinventariate, smarrite, scartate, ordini non evasi ecc.)

Altrettanto significative sono state le decisioni che hanno coinvolto gli LBD, sezione a uso esclusivo della biblioteca. Tutti i dati che non avrebbero potuto essere di interesse per la comunità di OCLC sono stati infatti mappati sui dati bibliografici locali, ma è stata necessaria un'attenta valutazione per scegliere in quali campi LBD, standard o proprietari, inserirli.

Questa valutazione ha coinvolto, per esempio, le note all'esemplare su record di fondi personali (tag MARC21 596, 597, 598) e la Classificazione decimale Dewey.

Quest'ultima decisione, che ha visto l'inserimento della CDD nel tag MARC21 650, è stata particolarmente faticosa, perché inizialmente c'era una forte incertezza se posizionare in un campo di soggetto il numero della classificazione. Alla fine, tale soluzione è stata adottata per gli indiscutibili vantaggi: i numeri CDD corredati di vedetta sarebbero stati visibili in discovery e soprattutto ricercabili.

Tra le scelte di carattere operativo si annoverano, invece, una serie di azioni finalizzate al miglioramento qualitativo dei dati e al loro allineamento a pratiche in uso nel contesto OCLC, come per esempio l'eliminazione delle parentesi uncinata (< >) nelle qualificazioni.

In seguito alla fase di analisi i record bibliografici sono stati inviati a OCLC in blocchi per la pre-elaborazione vera e propria: i raggruppamenti sono stati definiti in base alle caratteristiche dei record stessi, dal blocco più omogeneo (record di monografie libro moderno) a quelli più particolari (record descrittivi di opere in più volumi o di miscellanee non editoriali, oppure del libro antico, oltre naturalmente al blocco di record descrittivi dei seriali).

Gli interventi di pre-elaborazione e di sincronizzazione sono stati descritti in un documento chiamato *Scope statement* che ha avuto sei versioni successive, a prova sia dei ragionamenti e delle riflessioni che man mano emergevano, sia della buona collaborazione che c'è stata con i colleghi italiani e olandesi di OCLC.

Lo *Scope statement* ha raccolto le scelte da operare sia sui metadati bibliografici sia su quelli delle copie ed è stato organizzato per sezioni di lavorazione:

- descrizione di tutti gli interventi nella fase di pre-elaborazione;
- configurazione dei dati per la sincronizzazione;
- compilazione dei campi nella sezione LBD;
- mappatura dei metadati per i record di copia/LHR.

Conclusa la fase di pre-elaborazione, è iniziata la fase di caricamento dei dati, durante la quale OCLC ha avviato le procedure di importazione e *matching*, allo scopo di trovare le corrispondenze con i record già presenti in WorldCat e non creare duplicati. Gli elementi usati per individuare le corrispondenze includevano:

- numeri univoci come ISBN, ISSN o numero OCLC (OCN);
- tipo di materiale fisico;
- data di pubblicazione;
- lingua della catalogazione;
- titolo;
- autore;
- edizione;
- editore.

Il *matching* implicava tre diversi risultati: nessuna corrispondenza e creazione del nuovo record (*record added*), corrispondenza trovata con conseguente schiacciamento (*record merged*), corrispondenze multiple su più record. In presenza di più di una corrispondenza, veniva avviata una nuova procedura, denominata 'processo di risoluzione', allo scopo di scegliere il miglior record sulla base di alcuni algoritmi, tra cui:

- origine del record (campo MARC21 040 \$c);
- codici di autenticazione nel campo MARC21 042;
- livello di codifica e numero di localizzazioni.

A conclusione della fase di sincronizzazione, i record venivano inviati alla convalida.

Non per tutti i record bibliografici il processo si è concluso con la convalida e il conseguente caricamento e visualizzazione su WMS. Si parla in questo caso di *sparse record*, ovvero di record non convalidati a causa di specifici errori a livello di struttura del formato MARC21: gli *sparse record* sono stati comunque caricati, ma sono invisibili agli utenti e richiedono una correzione manuale. A tale scopo, OCLC ha fornito un'apposita reportistica sui record *sparse*.

Le sfide della migrazione tra cambiamenti e rinunce: alcuni casi

L'intero processo di migrazione è stato impegnativo: i record bibliografici trattati da OCLC sono stati 271.000. Non potendo scendere nel dettaglio di ogni decisione, fa seguito un breve approfondimento sulle due situazioni più complesse: la gestione delle opere in più volumi e i dati di copia. Le soluzioni finali adottate in questi due contesti hanno richiesto diversi momenti di riflessione.

Le opere in più volumi

Come si è detto in precedenza, la catalogazione a livelli era stata limitata, già prima della migrazione, a casi ben selezionati, preferendole soluzioni meno complesse, tra le quali l'uso della nota di contenuto.

Nel formato MARC21, la gestione dei legami verticali è garantita dal tag di relazione 773, corrispondente al tag UNIMARC 461; tuttavia, per una pratica interna definita dalla *community* di OCLC, il tag 773 non viene utilizzato a tale scopo all'interno di WMS.

Per questo motivo, gli unici record migrati con il legame 773 sono stati quelli del libro antico e delle miscellanee non editoriali: tuttavia, il campo di legame, privo

della citazione del numero di sistema, non ha creato un reale collegamento tra i record, ma fornisce solo l'indicazione del titolo superiore in forma testuale.

[Miscellanea Di Estratti Sulla Psichiatria]. 115. The Italian Psychiatric Reform: Some Quantitative Evidence

LEADER 00000cam#2200000#1#4500
ocalDataGroups ##
root ##
001 ## 1262081945
008 ## 210304c1987###gw#####000#0eng#d
040 ## \$a ITMBA \$b ita \$c ITMBA
245 00 \$a [Miscellanea di estratti sulla psichiatria]. \$n 115, \$p The Italian psychiatric reform: some quantitative evidence / \$c M. Tansella, D. De Salvia and P. Williams.
260 ## \$a Berlin [etc.] : \$b Springer, \$c 1987.
300 ## \$a 37-48 p. : \$c 25 cm
500 ## \$a Estratto da: Social Psychiatry, v. 22 (1987).
700 1# \$a De Salvia, Domenico.
700 1# \$a Tansella, Michele, \$d <1942-2015>
700 1# \$a Tansella, Michele, \$d (1942-2015)
700 1# \$a Williams, Paul, \$d (1952-)
773 0# \$t [Miscellanea di estratti sulla psichiatria]

Figura 5 – Opera in più volumi con tag 773

Il trattamento dei restanti record multivolume è dipeso invece dalla presenza o meno dell'opera in WMS. I *record merged*, confluiti in record già esistenti, presentano descrizioni di tipo diverso: con il titolo della superiore nella collana, con legame nel tag 773, con titolo composto, con nota di contenuto.

040 ## \$a ITCBT \$b ita \$c ITCBT \$d OCLCQ \$d ITCPE \$d OCLCQ \$d ITCBS \$d OCLCQ \$d ITMBA
041 0# \$a ita \$a eng
082 14 \$a 530
084 ## \$a ME \$2 msc
100 1# \$a Feynman, Richard Phillips. \$4 aut
245 13 \$a La fisica di Feynman / \$c Richard Philip Feynman.
246 0# \$a Feynman lectures on physics
260 ## \$a Bologna : \$b Zanichelli, \$c 2001.
300 ## \$a 3 v. : \$b illustrations ; \$c 20 x 27 cm
336 ## \$a testo \$b txt \$2 rdacontent/ita
337 ## \$a senza mediazione \$b n \$2 rdamedia/ita
500 ## \$a Testo anche in inglese.
505 0# \$a 2 : Elettromagnetismo e materia. -- 1 v. (varie sequenze).
505 0# \$a 3 : Meccanica quantistica. -- 1 v. (varie sequenze).
505 0# \$a 1 : Meccanica, radiazione, calore. -- 1 v. (varie sequenze).
650 #4 \$a Fisica.
650 #7 \$a Fisica. \$2 fi
653 0# \$a Fisica
700 1# \$a Leighton, Robert B. \$4 aut
700 1# \$a Sands, Matthew. \$4 aut

Figura 6 – Record merged con nota di contenuto

040 ## \$a ITCPE \$b ita \$c ITCPE \$d ITCPE \$d OCLCQ \$d UAV
041 0# \$a eng \$a ita
082 04 \$a 530 \$2 22
100 1# \$a Feynman, Richard P. \$4 aut
245 13 \$a La Fisica di Feynman. \$n 4, \$p Consigli per risolvere i problemi di fisica / \$c Richard P. Feynman, Michael A. Gottlieb, Ralph Leighton.
260 ## \$a Bologna : \$b Zanichelli, \$c 2007.
300 ## \$a XI, 145 p. ; \$c 21 x 28 cm
336 ## \$a testo \$b txt \$2 rdacontent/ita
337 ## \$a senza mediazione \$b n \$2 rdamedia/ita
653 0# \$a Fisica
700 1# \$a Leighton, Ralph. \$4 aut
700 1# \$a Gottlieb, Michael A. \$4 aut
765 0# \$t Feynman's tips on physics: a problem-solving supplement to the Feynman Lectures on physics.
773 0# \$t Fisica di Feynman \$g 4

Figura 7 – Record merged con titolo composto e tag 773

I *record added*, non presenti in WMS, sono stati migrati con la struttura del titolo composto. Per questo motivo, i volumi di una stessa opera possono avere descrizioni non omogenee.

LEADER 00000cam#2200000#1#4500
0ca1DataGroups ##
root ##
001 ## 1262090464
008 ## 190114z2000###ita######0000#0#ita#d
019 ## \$a 889195399 \$a 929680944 \$a 1261171420
020 ## \$a 8886797044
020 ## \$a 9788886797047
040 ## \$a IIMBA \$b ita \$c IIMBA \$d ITCPE
082 04 \$a 945.24 \$2 23
130 3# \$a Il tempo della Serenissima. Un Seicento in controtendenza.
245 10 \$a Storia economica e sociale di Bergamo. \$n 3,3, \$p Il tempo della Serenissima. \$p Un Seicento in controtendenza / \$c [a cura di Aldo De Maddalena, Marzio Achille Romani, Marco Cattini].
260 ## \$a Bergamo : \$b Fondazione per la storia economica e sociale di Bergamo, Istituto di studi e ricerche, \$c 2000.
300 ## \$a 293 p. : \$b ill. ; \$c 27 cm
336 ## \$a testo \$b txt \$2 rdacontent/ita
337 ## \$a senza mediazione \$b n \$2 rdamedia/ita
653 0# \$a Bergamo \$a Storia \$a Sec. 17.
700 1# \$a De Maddalena, Aldo.
700 1# \$a Cattini, Marco.
700 1# \$a Romani, Marzio A.
773 0# \$t Storia economica e sociale di Bergamo

Figura 8 – Record added con titolo composto

La gestione delle copie

Analogamente ai record bibliografici, anche i record di copia presentano delle significative differenze rispetto al precedente gestionale.

In primo luogo, in WMS i dati di possesso si articolano su due livelli: la *holding* (il posseduto della biblioteca) e il *local holdings record* (LHR), cioè i dati di collocazione. Ciò significa che un record bibliografico può essere contrassegnato come posseduto ma essere privo di LHR.

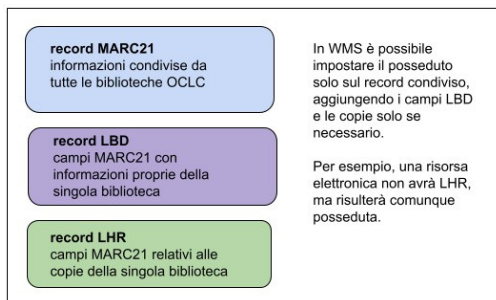


Figura 9 – Struttura tripartita del record in WMS

Il LHR in WMS può essere visualizzato in una duplice modalità: in visualizzazione testo e in formato MARC21. Aleph utilizzava l’UNIMARC per il record bibliografico e il MARC21 per il record di copia, ma offriva la visualizzazione dei dati di copia solamente tramite delle etichette parlanti, rendendo l’utilizzo dei campi MARC21 invisibile ai bibliotecari. Il passaggio a WMS ha quindi comportato l’apprendimento dei campi MARC21 delle *holding*, in aggiunta ai campi MARC21 bibliografici.

Si è resa necessaria un’apposita mappatura per i dati di copia (giugno-settembre 2021), di cui si riassumono di seguito alcuni aspetti significativi:

- solo il tag 852, con i dati di collocazione, è migrato automaticamente; tutti gli altri sono stati mappati singolarmente;
- le informazioni relative allo stato della copia (prestabile, consultazione ecc.) e all’indicazione del fondo (ad esempio collezione generale, testo ecc.), sono entrambe confluite nella cosiddetta *shelving location*. Nella *shelving location* i diritti di prestito sono concettualmente sempre legati alla collocazione fisica e viceversa. Presso la Biblioteca di ateneo, però, i libri sono raggruppati secondo altri criteri, e copie con prestabilità diverse coesistono sullo stesso ripiano. Questo ha richiesto la creazione di una *shelving location* per ogni combinazione di collocazione fisica e prestabilità (per esempio, la sede di Scienze e la sede di Medicina hanno entrambe la propria *shelving location* “Scaffale aperto - prestabile”). Grazie alla presenza di etichette uguali per ogni sede, questa struttura complessa rimane comunque invisibile agli utenti.

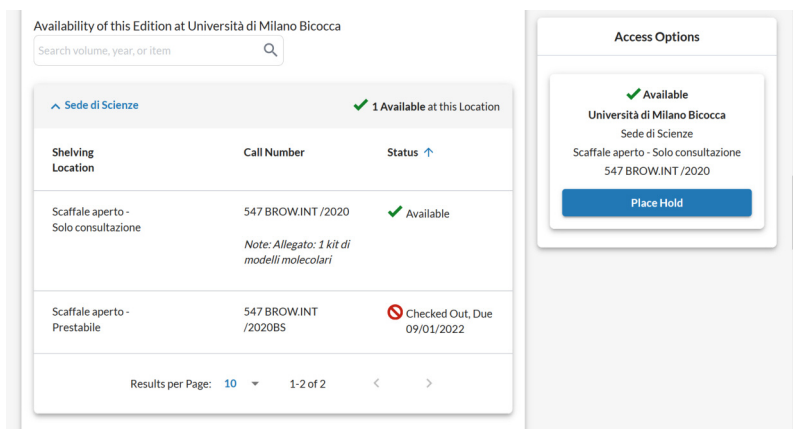


Figura 10 – Presenza di *shelving location* diverse per copie nella stessa sede

Le informazioni sullo stato di processo (ad esempio in catalogazione, smarrito, scartato ecc.) sono state assorbite in parte dalla *shelving location* e in parte dalle condizioni dell'ordine o dai dati della circolazione. Per esempio, "In legatoria" è diventata una *shelving location*.

Conclusioni

La transizione da Aleph a WMS è stata piena di sfide, prima fra tutte l'adattarsi a un nuovo formato di catalogazione.

Il passaggio dal formato UNIMARC al formato MARC21 ha implicato un ripensamento notevole nella formulazione e articolazione di specifici dati e nell'approccio ad alcuni aspetti catalografici di primaria importanza. Per fare un esempio, la maggiore visibilità data all'autore principale nel formato MARC21 oppure la differente strutturazione dei campi per i titoli uniformi.

Se il passaggio a un nuovo formato è, di per sé, attività non facile né immediata, la contestualizzazione all'interno di un sistema complesso e con regole proprie come WMS ha richiesto un ulteriore 'salto': l'adeguamento del formato si è associato all'apprendimento di pratiche interne a cui è stato necessario uniformarsi. Ne sono un esempio il trattamento dei multivolume – che in WMS differisce dallo standard MARC21 – e la composizione del campo riservato alla Classificazione decimale Dewey. Il passaggio alla catalogazione condivisa ha quindi comportato la necessità di rinunciare al controllo sui dati offerto da un ambiente locale.

Allo stesso tempo, la vastità del database consente di trovare spesso record già catalogati, che è sufficiente localizzare o derivare, con un risparmio di tempo e di risorse; questo vale anche per le risorse elettroniche, che per la prima volta sono totalmente integrate in unico *workflow* di catalogazione e di acquisizione con le risorse a stampa.

Inoltre, la partecipazione a una comunità globale offre la possibilità di confrontarsi con i colleghi e di contribuire allo sviluppo dello strumento attraverso richieste di *enhancement* condivise.

Non ultimo, la presenza del nostro posseduto all'interno di WorldCat ci offre una grande visibilità a livello globale.

Dal punto di vista dell'utente, invece, il passaggio a WMS e a WorldCat Discovery si è tradotto in una migliore disponibilità dei vari formati di una risorsa (a stampa e in elettronico) e nella possibilità di estendere agevolmente la ricerca all'intero posseduto di WorldCat.

La situazione attuale compensa quindi ampiamente il lavoro richiesto dal cambio di formato.

Articolo proposto il 18 luglio 2022 e accettato il 6 settembre 2022.

ABSTRACT AIB studi, vol. 62 n. 3 (settembre/dicembre 2022), p. 519-533. DOI 10.2426/aibstudi-10935
ISSN: 2280-9112, E-ISSN: 2239-6152 - Copyright © 2022 Alessandra Moi, Silvana Re,
Francesca Verga

ALESSANDRA MOI, Università degli studi di Milano-Bicocca, e-mail alessandra.moi@unimib.it.

SILVANA RE, Università degli studi di Milano-Bicocca, e-mail silvana.re@unimib.it.

FRANCESCA VERGA, Università degli studi di Milano-Bicocca, e-mail francesca.verga@unimib.it.

«Dietro è la casa, davanti a noi il mondo»: da UNIMARC a MARC21 all'Università di Milano-Bicocca
Il sistema gestionale è il fulcro di tutte le attività di una biblioteca, dalla catalogazione alle acquisizioni alla circolazione, e la mancanza di un modulo per gestire le risorse elettroniche penalizza le attività

quotidiane. La scelta di una nuova piattaforma integrata, però, può comportare sia vantaggi che sfide, tra cui il passaggio a un nuovo formato di catalogazione. Nel 2021, la Biblioteca di ateneo dell'Università di Milano-Bicocca ha scelto di adottare il gestionale Worldshare Management Services (WMS) di OCLC, affiancando al cambio di software il passaggio da UNIMARC a MARC21. L'articolo vuole condividere le scelte e gli aspetti cruciali della transizione, soprattutto da un punto di vista catalogafico: in particolare, si sofferma sul passaggio da un ambiente locale a uno condiviso, e sull'influenza che il nuovo formato adottato ha avuto sulle scelte catalogafiche.

«Home is behind, the world ahead»: from UNIMARC to MARC21 at the University of Milano-Bicocca

A library management system is the hub of all library activities, from cataloging, to acquisitions, to circulation, and the lack of an ERM (Electronic Resource Module) takes a heavy toll on day-to-day work. On the other hand, the choice of a new LSP (Library Service Platform) may present both advantages and challenges, such as switching to a different cataloging format. In 2021, the Milano-Bicocca University Library chose to adopt the Worldshare Management Services (WMS) LSP by OCLC, which meant facing a migration from a different management system, and also from UNIMARC to MARC21. The aim of this article is to share the pivotal decisions and moments in this transition from a cataloger's point of view: in particular, it focuses on moving from a local to a shared environment, and on the ways the new format influenced cataloging choices.