

On line union catalogue (Olc) compie 25 anni

Origine e sviluppo della rete Oclc e del suo famoso catalogo cooperativo

di Gabriele Lunati

The Rand McNally book of favorite pastimes è un piccolo libro (14 cm.) illustrato, per bambini, che probabilmente non varrebbe la pena di citare se non avesse la particolarità di essere il record n.1 di Oluc, la base dati bibliografica di Oclc avviata nel 1969, ufficialmente aperta nel 1971 e che al momento in cui scrivo è giunta al record n. 34.099.794.

Olc sta per On line union catalogue il catalogo della rete Oclc (On line computer library center) che, reso disponibile alle biblioteche dell'Ohio College nel 1971, ha appena compiuto il suo venticinquesimo anno di attività.

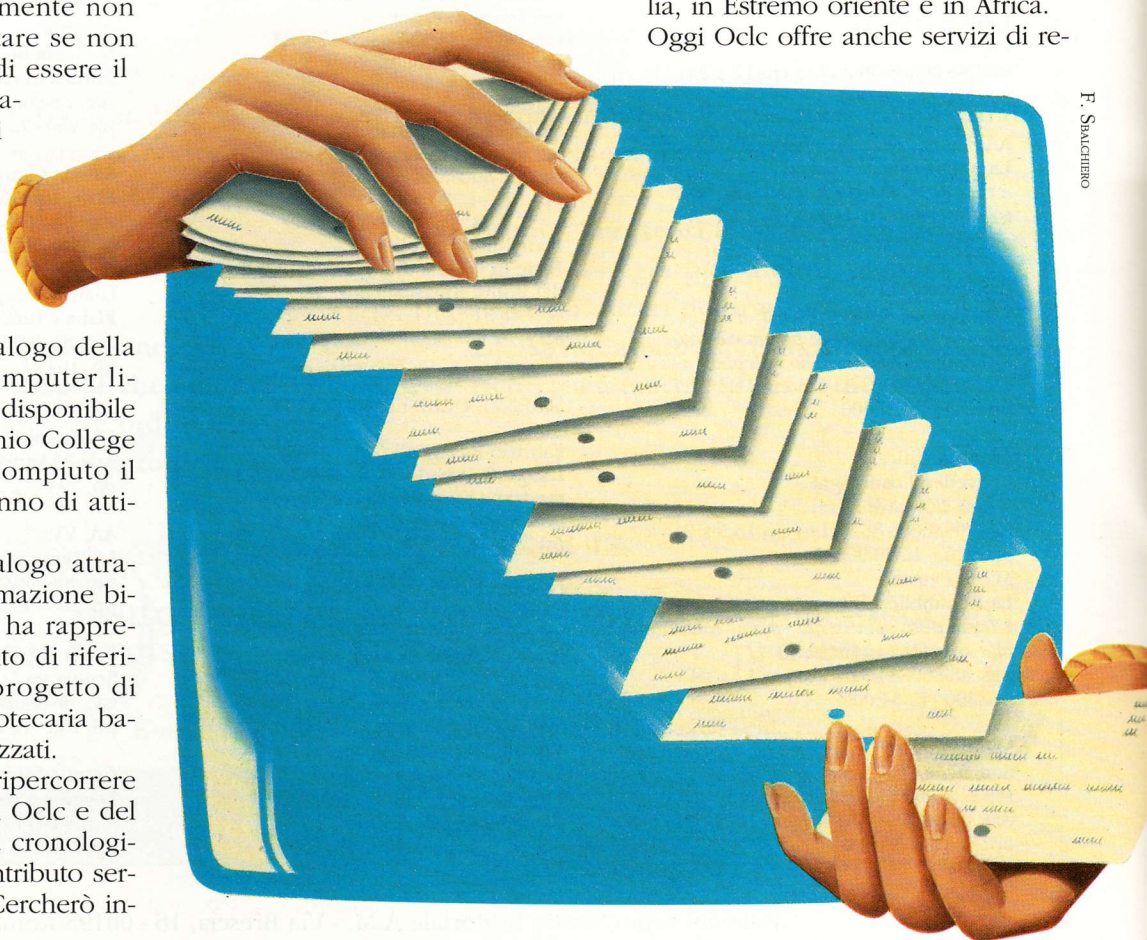
La storia di questo catalogo attraversa la storia dell'automazione bibliotecaria nel mondo, ha rappresentato sempre un punto di riferimento per qualsiasi progetto di cooperazione interbibliotecaria basato su sistemi automatizzati.

Sarebbe troppo lungo ripercorrere nel dettaglio la storia di Oclc e del suo catalogo; la scheda cronologica inclusa in questo contributo servirà come riferimento. Cercherò in-

vece di fissare alcuni degli aspetti più importanti di questa impresa cooperativa, puntando lo sguardo

piuttosto sugli investimenti e sugli sforzi organizzativi che con continuità sono stati fatti per mantenere adeguata questa grande risorsa.

L'originaria richiesta fatta a Frederick Kilgour ed al suo staff di dotare le biblioteche dell'Università dell'Ohio College di Columbus di un sistema automatizzato di circolazione dei documenti, si trasformò ben presto in un progetto di condivisione delle risorse ben più ampio e più organico che, per forza di cose, una volta avviato e sperimentato con successo all'interno del campus originario, avrebbe finito per interessare altre università e poi altre biblioteche anche non universitarie della stessa regione e poi della stessa nazione. Cosa avvenuta puntualmente fino a superare i confini sia dell'Atlantico che del Pacifico e approdare nell'arco di 10-15 anni in Europa, in Australia, in Estremo oriente e in Africa. Oggi Oclc offre anche servizi di re-



F. SANCHEIRO

ference come FirstSearch, prodotti su cd-rom, soluzioni informatiche avanzate come SiteSearch, ha acquisito la Forest Press e distribuisce la classificazione Dewey sia a stampa che in forma elettronica, si è lanciato nella produzione di riviste elettroniche, ha incamerato importanti aziende come la Information Dimension, attraverso cui lancia un ponte anche in direzione della gestione automatizzata delle biblioteche. Ma l'oggetto più importante e più vitale della sua attività era e resta Oluc, il catalogo cooperativo, lo strumento su cui Oclc orienta i suoi sforzi maggiori e con il quale vuole vincere la scommessa della "qualità totale" entro l'avvio del secondo millennio.

Nel corso di questi 25 anni molte cose sono cambiate con una particolare velocità durante gli ultimi cinque anni quando più forte si è fatta la concorrenza, soprattutto nel settore delicato ed immenso della riconversione dei cataloghi e della catalogazione derivata. Il nucleo delle biblioteche originarie, ampio ma prevalentemente americano, ha fatto posto a numerosi partner di altri paesi, sia in Europa che in altri continenti. L'impatto con molte biblioteche europee, in particolare della Francia e del Regno Unito, ha posto ad Oclc dei problemi di correttezza e completezza dell'informazione bibliografica e di aderenza a determinati standard qualitativi che mai prima li aveva toccati così drammaticamente.

Sono del 1993 il grande progetto di riconversione avviato con la Harvard University e l'accordo, con successivo riversamento dei dati, con la rete universitaria inglese Curl (Consortium of university and research libraries). Due avvenimenti di grande portata politica e bibliografica, di fronte ai quali il fatto di avere già avviato una serie di iniziative pratiche a tutela della qualità dei dati, fu particolarmente importante.

Le critiche sulla "correttezza" e sulla "unicità" dei record si erano fatte pressanti, il confronto con la più selettiva e scientifica Rlin sempre più fastidioso.¹

Oluc era il catalogo molto ricco ma molto sporco, quindi potenzialmente meno ricco e meno affidabile. Negli anni Settanta era aneddoto comune citare il fatto che su Oluc non si pagava se si inseriva un nuovo record e così tutti si sbizzarivano a cambiare qualcosa in modo da risparmiare, perdendo però tutti alla fine in qualità.²

L'impatto negativo maggiore era quello sul servizio di prestito interbibliotecario (Ill), fiore all'occhiello della rete, per funzionalità e semplicità. L'efficienza di Prism-Ill (catalogazione e prestiti unificati) rischiava di essere vanificata disperdendo gli sforzi dei bibliotecari nella ricerca ed acquisizione dei documenti, tra record duplicati ed errori di varia natura.

La riflessione degli utenti e le conseguenti misure non si fecero attendere. Oclc, vale la pena di ricordarlo, è una organizzazione *non profit*, cioè senza fini di lucro, in cui l'utente attraverso una serie molto articolata di organismi consultivi e decisionali, delibera su tutti gli aspetti tecnici e scientifici. Manager, non bibliotecari ma provenienti da grandi industrie, sintonizzano le decisioni prese anche alle logiche di mercato facendo sì che Oclc sia impresa in utile costante e in grado di reinvestire.

Così, all'inizio degli anni Novanta in concomitanza con la grande espansione internazionale, sono stati avviati i programmi per la cura della qualità.

Non bastava potenziare le ricerche, aprire il servizio 23 o 24 ore al giorno, proporre nuovi strumenti e nuove chiavi di accesso alle oramai logore, anche se sempre utilizzabili "chiavi derivate".³ Bisognava fare opera di sensibilizzazione alla correttezza e al recupero dell'informa-

zione, attivare degli strumenti di controllo automatici, se e dove possibile, oppure forme di cooperazione che si orientassero a responsabilizzare determinati gruppi o tipi di biblioteche per garantire un controllo bibliografico di livello qualitativo elevato. I risultati non sono mancati. L'utenza internazionale ha continuato ad espandersi e le prestazioni della rete sono cresciute in proporzione.⁴

Alla fine del 1993 Oclc ha pubblicato un documento programmatico molto impegnativo dal titolo *Oclc's Cataloguing and database services and strategies*. Si tratta di uno studio redatto dal centro studi di Oclc che disegna lo scenario degli anni Novanta per definire le linee di condotta del quinquennio finale del secolo, non solo prospettando il massimo della qualità ma anche proponendo nuovi servizi per istituti che, di anno in anno, vedevano ridursi le proprie risorse economiche. Lo scenario di riferimento era quello di biblioteche con budget ridotti, costi di gestione elevati, necessità di limitare i servizi e problemi di personale; tra tutte le operazioni la più costosa: la catalogazione.

Oclc nel triennio 1990-1993 aveva avviato grandi cambiamenti: migliorato decisamente le potenzialità della ricerca, potenziato le funzioni di editing e di validazione dei record bibliografici con un primo tentativo concreto di minimizzare gli errori, introdotto procedure più amichevoli di uso del sistema.

Quali gli obiettivi degli anni a venire fino al 2000: il miglioramento della produttività a minor costo, l'attivazione di servizi di authority control e l'arricchimento della base dati Oluc.

Bisognava anche introdurre servizi che collegassero la catalogazione alle procedure di acquisizione, modalità di classificazione facilitata dei documenti e servizi di catalogazione "chiavi in mano". ➤

I risultati di queste decisioni sono due servizi avviati in questi ultimi due anni che si chiamano Techpro (la catalogazione "chiavi in mano") e Prompt-Cat (per gli acquisti).

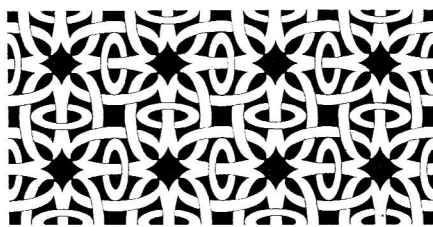
Se il primo interessa prevalentemente, anche per ragioni geografiche, l'area statunitense, il secondo, Prompt-Cat, ha cominciato a interessare anche l'Europa. Casalini libri per l'Italia e Puvill Libros per la Spagna sono stati in Europa i primi aderenti a questo sistema che, sulla base dei dati bibliografici delle novità, permette alle biblioteche di avviare procedure di acquisizione e di catalogazione derivata.

Altri punti determinanti indicati dallo studio erano l'apertura verso i sistemi locali e la disponibilità di tutti i servizi su Internet. Per il primo obiettivo è stato realizzato un software di interfaccia denominato Gateway per connettere stazioni su una rete locale e terminali non Oclc ai servizi di Oclc; i servizi sono stati resi compatibili con lo standard Z39.50 e la connessione via Internet è stata resa possibile a partire dall'aprile 1994. Ma il punto più impegnativo, su cui Oclc ha dovuto concentrarsi e orientare i propri investimenti è stato il problema della qualità dei dati.

Nel momento in cui si parla di aumentare la produttività individuale, l'esigenza di effettuare minori controlli, confronti e correzioni diventa fondamentale. La ricerca può così mantenersi su alti standard di definizione e la correttezza del dato ha ricadute positive sulle procedure gestionali. Le questioni di Oclc da risolvere erano concentrate in tre "motti": "Too much / too little / too many".

Il problema dei record duplicati (*too much*) era certamente il più grosso e necessitava di essere affrontato con decisione prima che la massa critica della base dati rendesse le operazioni di deduplicazione troppo complesse. Il secondo problema era quello della po-

vertà dei dati (*too little*). Record bibliografici introdotti a livello minimo e sufficienti per la gestione locale, non potevano alla lunga essere accettabili per la catalogazione derivata e i progetti di riconversione, comportando procedure di completamento ed arricchimento troppo costose. Infine la ridondanza delle intestazioni sia autore che soggetto (*too many*) denotava una scarsa dimestichezza con gli authority file e con l'applicazione degli standard. A tutto questo andavano aggiunti gli errori di codifica Marc, non meno frequenti di altri.



L'atteggiamento dei catalogatori su Oclc talvolta lasciava perplessi. La loro adesione agli standard catalografici pareva fine a se stessa piuttosto che orientata ad arricchire il catalogo di informazioni significative. Così Walter High in un breve studio pubblicato nel 1991, dove, fra l'altro, osservava che: "più della metà degli interventi sui record rispondeva ad esigenze estetiche (cosmetic) più che ad incrementarne la utilizzabilità".⁵ C'era dunque anche un problema culturale.

La risposta di Oclc per arrivare ad una soluzione di questi problemi è stata molto articolata e orientata in varie direzioni simultaneamente.

Accanto alla soluzione dei controlli manuali, da sempre presenti, sono stati attivati automatismi via software e si è giocata ancora una volta, e in modo esteso, la carta della cooperazione e della responsabilizzazione delle biblioteche. La modalità originaria di controllo dei duplicati ha preso avvio dai primi anni Settanta e si è basata sul meccanismo della segnalazione da par-

te dei membri di Oclc e sul lavoro di una apposita commissione (Odqcs: Oclc duplicates & quality control staff). Fino al 1991 sono pervenute ad Oclc circa 23.000 richieste di correzione e il gruppo Odqcs ha per conto proprio deduplicato oltre 267.000 record. Un'impresa titanica tenuto conto della crescita media della base valutabile intorno ai 2-2,5 milioni di nuovi record all'anno.

Lo studio di soluzioni automatizzate, partito dal 1986, ha generato tre distinti programmi denominati Ddr per i libri, Sema per i periodici e Aacs per le intestazioni. Si tratta in ogni caso di algoritmi di ricerca che hanno lo scopo di verificare elementi convergenti tra record diversi,⁶ attribuendo un "peso" al record; questo parametro permette, quando supera una certa soglia, di definire i record come doppi o potenzialmente doppi, effettuando una operazione di "fusione" o isolandoli per ulteriori analisi. Reso operativo nel 1991 il Ddr (Duplicate detection & resolution) ha prodotto in tre anni il corrispettivo di un trentennio di lavoro manuale, deduplicando oltre 700.000 record: il software Sema (Serial extended matching algorithm) avviato nel 1992 si è reso particolarmente utile in occasione di massicci caricamenti via nastro forniti dai nuovi importanti aderenti alla rete degli ultimi anni.

Più delicato il compito del terzo algoritmo, quello dedicato alle intestazioni. L'Aacs (Automated authority control software) è stato concepito come software intelligente, basato sulle Aacr2 (il codice di catalogazione anglo-americano) e destinato ad eliminare da Oclc errori di stile, tipografici, abbreviazioni, termini obsoleti e varianti non più accettate. Nell'arco di un biennio gli effetti si sono fatti sentire con oltre un milione di intestazioni autore corrette e quasi due milioni di soggetti.

Tab. 1 - Esempi di dati modificati

Errore	da	a
stile	KLINGER, Hans Herbert	KLINGER, Hans-Herbert
tipografico	centeral nervous ...	central nervous ...
abbreviazione	govt.	government
termine obsoleto	Edinburgh (Lothian)	Edinburgh (Scotland)
variante	KLAIBER, Teresa	KLAIBER, Teresa Lynn Martin 1949-

Nei primi mesi del 1996 Oclc ha inaugurato un nuovo servizio: l'Authority control service, in grado di fornire forme controllate per forme autori personali, enti, serie, soggetti sia della Library of Congress (Lcsh) che di argomento medico (Mesh).

Per sottolineare l'importanza di questo lavoro basti pensare che alla fine del 1994 Oluc forniva informazioni su circa 29 milioni di record e che le intestazioni erano complessivamente quasi 3,5 milioni. Ovviamente, un punto importante era garantire che il lavoro di verifica e correzione non diventasse una specie di tela di Penelope, vanificandone grazie a nuove imprecisioni e scorrettezze i risultati raggiunti. Sono state perciò introdotte nuove funzionalità di validazione e controllo sui nuovi inserimenti, per eliminare in primo luogo errori di uso del formato Marc e poi sono stati avviati alcuni importanti programmi di cooperazione per garantire record più corretti e completi.

Due di questi programmi riguardano i periodici. Si tratta di Conser (Cooperative online serials) ed Usnp (United States newspaper program). Conser, a cui partecipano una ventina di biblioteche, è di interesse generale, orientato alla normalizzazione dei dati sui periodici e i risultati del lavoro sono disponibili dalla metà di quest'anno anche su

un cd-rom oltre che sulla rete.

Usnp è invece orientato alla catalogazione del posseduto e delle consistenze, grazie alla creazione e manutenzione delle cosiddette Union List; fondamentale strumento per la circolazione dei documenti, utile per sapere non solo chi possiede un certo periodico, ma se ha proprio l'annata che serve.

Per la cura degli authority file è operativo il Naco (National coordinated cataloguing operation) a cui partecipano una cinquantina di biblioteche in collaborazione con la Library of Congress. Ben tre i programmi cooperativi destinati all'arricchimento della base Oluc.

Al programma Enhance (che una volta tanto non è una sigla) collaborano oltre un centinaio di biblioteche responsabili della correzione di determinate aree della base e da qualche anno investite anche del completamento delle catalogazioni Cip (Cataloguing in publication).

Il Nccp (National coordination cataloguing program) vede all'opera otto biblioteche la cui autorità catalografica è stata elevata al livello della Library of Congress e che sono responsabili di produrre descrizioni ad alto livello.⁷

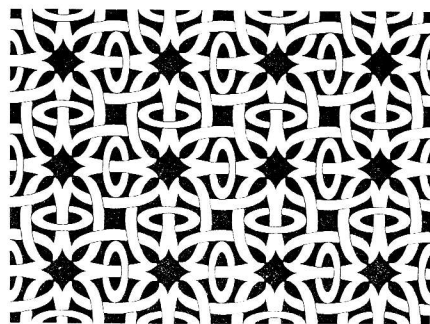
Infine il Fiction project, dedicato non solo alla descrizione ma anche alla soggettazione di opere di narrativa. Vi partecipano cinque biblioteche coordinate dalla onnipotente Library of Congress.⁸

Conclusioni

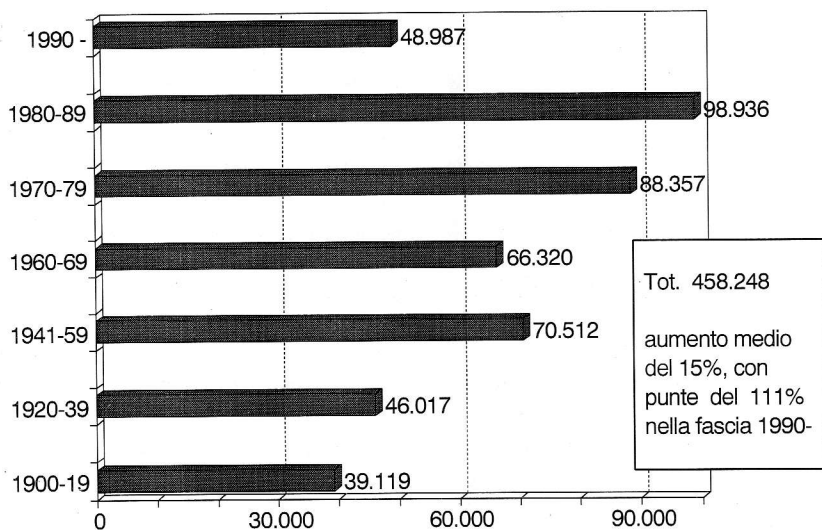
Meriterebbe maggiore spazio ciascuno progetto, la metodologia di lavoro impiegata e l'analisi dei problemi incontrati. Non mi pare il caso di infierire in questa sede, considerato che l'assunto di partenza era quello di illustrare come, con quali tempi e risultati Oclc ha affrontato il problema della qualità.

Le ricadute di questo lavoro, considerata l'espansione a livello mondiale di questa rete, saranno importanti a qualsiasi latitudine, anche in presenza di standard catalografici differenti. Interessa dunque le biblioteche europee e, perché no, quelle italiane.

In Italia Oclc ha iniziato ad essere utilizzato con crescente attenzione a partire dal 1994. Prevalentemente, all'inizio, per i prodotti su cd-rom, poi per il prestito interbibliotecario (Prism-III) e con il 1995 in modo significativo anche per la catalogazione derivata. La pregiudiziale nei suoi confronti in quanto base dati "angloamericana" si sta stemperando grazie ai risultati che vengono ottenuti; la sensazione che Oluc sia povero di pubblicazioni italiane è smentita dai fatti (si vedano in proposito le tabelle 2 e 3 nella pagina seguente); l'idea che offra dati "sporchi" dovrebbe infine rientrare alla luce di quanto Oclc ha già realizzato per il raggiungimento di alti standard qualitativi.⁹ In ogni caso, se non altro a livello metodologico, la strada percorsa da Oluc in questi suoi primi 25 anni di vita, merita considerazione. ►



Tab. 2 - Pubblicazioni italiane del Novecento presenti nella base dati Oluc (dati rilevati il 9/6/96 su 607.028 record)



Note

¹ Rlin (Research library information network) la seconda rete interbibliotecaria mondiale per importanza, è il concorrente storico di Oclc. Condivide, tuttavia, con Oclc oltre la metà delle proprie fonti catalografiche; tra esse la Library of Congress.

² Oluc comprende pubblicazioni in circa 400 lingue (l'italiano rappresenta il 2 per cento) e copre praticamente qualsiasi tipo di materiale documentario dai libri ai file elettronici. L'86 per cento sono libri e il 5 per cento periodici. Oltre 600 milioni le localizzazioni agganciate ai record.

³ Le chiavi derivate previste da Oclc sono tre: titolo (3,2,2,1), autore (4,3,1) e autore/titolo (4,4). Dal 1993 sono state rese possibili ricerche per editore, luogo, anno, lingua, soggetto, termine del titolo: è stato introdotto l'uso dell'operatore booleano "and".

⁴ Sono circa 8.000 le biblioteche concorrenti alla base dati, con 125.000 catalogazioni giornaliere, circa 65 interrogazioni di media al secondo su un sistema tarato a sostenerne fino ad 85 e con tempi medi di risposta inferiori ai 4 secondi. Oltre 21.000 sono in tutto le biblioteche che usufruiscono dei servizi di Oclc.

⁵ W. HIGH, *How catalogers really edit Oclc records*, "North Carolina Libraries",

49 (Fall '91), p. 163-164.

⁶ L'algoritmo arriva a confrontare fino a 14 diversi elementi presenti in un record; tra essi per esempio: il titolo, l'Isbn /Issn, il luogo di pubblicazione.

⁷ Alla fine di Aprile 1996 la Library of

Congress, in collaborazione con Oclc, ha avviato una sperimentazione semestrale per la creazione di cosiddetti "core level bibliographic records", cioè di registrazioni bibliografiche un po' meno ricche del massimo ("full level") ma molto più ricche del livello minimo. Lo scopo è quello di confrontare costi e produttività rispetto al livello massimo. Maggiori informazioni sullo standard del "core level" sono disponibili alla pagina Web: <http://lcweb.loc.gov/catdir/pcc/pcc.html>.

⁸ Una lista aggiornata al 1993 delle biblioteche partecipanti ai vari progetti è stata pubblicata nel n. 204 (July/August 1993) di "Oclc Newsletter".

⁹ Un test incrociato, che è stato effettuato lo scorso anno scambiandosi i dati tra Oclc e un gruppo di biblioteche tedesche — altro osso duro in Europa — ha dato risultati sorprendenti, non solo rispetto alla copertura, ma anche rispetto alla percentuale di interventi di adattamento necessari per i tedeschi. Circa il 70 per cento dei record individuati andava bene così come si presentava su Oclc. Un analogo test si svolgerà nell'ultimo trimestre del 1996 tra un gruppo di biblioteche italiane e Oclc, sotto il coordinamento dell'Iccu ed in collaborazione con l'agenzia italiana di Oclc.

Tab. 3 - Pubblicazioni italiane del pre-Novecento presenti nella base dati Oluc (dati rilevati il 9/6/96 su 607.028 record)

