

Conservando più copie rendiamo sicure le nostre collezioni digitali

Gianluca D'Amato
Renato Mazzocchi

*La sperimentazione della tecnologia LOCKSS¹
presso il Sistema bibliotecario
dell'Università Cattolica del Sacro Cuore*

*Università Cattolica del Sacro Cuore, Milano
gianluca.damato@unicatt.it
renato.mazzocchi@unicatt.it*

Premessa

L'avvento dell'editoria elettronica distribuita via Internet, quale nuovo importante paradigma editoriale sia per i periodici scientifico-accademici sia per le fonti informative di ricerca primaria nell'ambito delle scienze umane e sociali (ad esempio, i giornali su Internet, i siti web di informazione o di discussione ecc.), rappresenta per le biblioteche una notevole opportunità di migliorare e potenziare i servizi offerti agli *utenti attuali*. Allo stesso tempo, però, nella prospettiva dei servizi che la biblioteca intende prestare agli *utenti futuri* non si possono trascurare le criticità presentate dal formato elettronico ai fini della conservazione a lungo termine.

Gli interrogativi posti da questo nuovo scenario non hanno potuto non trovare ascolto nelle valutazioni che interessano la funzione acquisizione presso la Biblioteca dell'Università Cattolica del Sacro Cuore (funzione che nel nostro contesto prescinde dalla circostanza che l'accessione avvenga a titolo di acquisto oneroso, di cambio o di omaggio). Al netto di altre considerazioni (ad es. la questione riguardante l'IVA), si è venuta prospettando una alternativa radicale. Da un lato, l'acquisizione dell'ac-

cesso al formato elettronico avrebbe certamente accontentato gli utenti presenti, in termini di comodità e rapidità di accesso ai contenuti, con il contestuale vantaggio, per la biblioteca, di non aggravare i problemi di spazio. Dall'altro lato, l'acquisizione della proprietà delle edizioni a stampa, sulla cui conservazione a lungo termine si dispone di una esperienza assai ben consolidata, avrebbe tutelato gli utenti futuri, pur al prezzo di qualche scomodità per gli utenti attuali.

La nostra biblioteca ha risposto a questa problematica individuando, in una prima fase, una soluzione di medio periodo. Una biblioteca universitaria non può venir meno alla sua ragion d'essere, che è quella di mettere a disposizione materiale informativo per la ricerca e la didattica, tutelando sia gli utenti presenti sia quelli futuri. Ciò significava, in concreto, continuare ad acquistare l'edizione a stampa come edizione di riferimento, che sapevamo come conservare a regola d'arte, e considerare l'accesso all'edizione elettronica come un servizio aggiuntivo, da acquisire solo nei casi in cui la spesa aggiuntiva corrispondente fosse di importo modesto. Per quanto riguarda le pubblicazioni accessionate in omaggio, si decise di continuare a conservare la copia a stampa e da-

re notizia agli utenti della disponibilità dell'accesso gratuito all'edizione web.

Due elementi hanno contribuito più di altri a rendere evidente la necessità di affiancare, alla soluzione ora descritta, una soluzione di più lungo periodo: il costo crescente dell'accesso all'edizione elettronica, divenuto ora tale da non poter più essere sostenuto dal bilancio della biblioteca come servizio aggiuntivo, e la diffusione di pubblicazioni, di rilevante interesse accademico e/o di ricerca, disponibili nella sola edizione elettronica (si pensi, ad esempio, a Berkeley Electronic Press).

Mossi dalla convinzione che la biblioteca dovesse continuare a svolgere il ruolo che le è proprio anche di fronte al diffondersi di questo modo nuovo e rivoluzionario di fare editoria, abbiamo incominciato a studiare le diverse soluzioni disponibili nel panorama internazionale. Le questioni centrali a cui intendevamo trovare una risposta il più possibile adeguata si potrebbero formulare così:

- Soddisfatte quali condizioni il Sistema bibliotecario dell'Università Cattolica sarebbe stato in grado di *sostituire* edizioni a stampa accessionate in omaggio o in acquisto con edizioni in formato elettronico?
- Come preservare pubblicazioni

free o a pagamento disponibili solo in edizione elettronica?

Posti di fronte a questi interrogativi sostanziali, abbiamo cercato di informare la nostra azione a un duplice atteggiamento. Da un lato, la *prudenza*: non bisogna dimenticare che la storia degli ultimi secoli delle biblioteche è stata tutto sommato una storia di successo, almeno per il mondo occidentale, tanto che normalmente si riesce ad accedere a contenuti pubblicati nel Settecento o nell'Ottocento. Da questo punto di vista, la sfida per i bibliotecari di oggi è che il loro lavoro non sia da meno rispetto a quello svolto con le edizioni a stampa dai colleghi che li hanno preceduti.

D'altra parte, fatta salva la prudenza, abbiamo voluto unire a essa l'*efficacia*: non avere troppa fretta di decidere, ma trovare soluzioni vere, capaci di reggere in una prospettiva di lungo periodo, nella consapevolezza che alle sfide poste dall'editoria scientifica distribuita via web si risponde con un paniere di soluzioni e non con una panacea.

La conclusione a cui siamo in tal modo giunti è che, per un passaggio consapevole dalle edizioni a stampa alle edizioni elettroniche, occorre che siano soddisfatte le seguenti condizioni:

– La proprietà dei dati, o in subordine la loro piena disponibilità, su hardware e software di proprietà, da parte del personale della nostra biblioteca. In altri termini, la biblioteca tende a cercare nei supporti elettronici, capaci di conservare le informazioni, garanzie simili a quelle offerte oggi dal deposito e dagli scaffali in genere per le edizioni a stampa. Come condizione alternativa, la proprietà dei dati, o in subordine la loro piena disponibilità, su hardware e software di proprietà, da parte di enti non profit (per es. i consorzi), eventualmente nati dalla cooperazione tra biblioteche omogenee, rispetto ai quali il sistema bibliote-

cario della nostra università mantenga un ruolo importante di indirizzo e di controllo.

– La possibilità di realizzare copie di backup di sicurezza su server di proprietà UCSC (Università Cattolica del Sacro Cuore) o di proprietà di enti non profit che abbiano le caratteristiche di cui al punto precedente.

– L'impiego di formati elettronici che, sulla base delle conoscenze attuali, siano aggiornabili e trasferibili nel corso degli anni con sufficienti garanzie di accessibilità e di sicurezza.

– Aggiornamento software e hardware con costi sostenibili nel tempo. Tenendo presenti queste condizioni, abbiamo cominciato a prendere in esame le esperienze e i risultati già conseguiti da altre biblioteche e, in particolare, a documentarci su come talune biblioteche operanti all'estero, con una dotazione di mezzi superiore alla nostra, avevano affrontato il nostro stesso problema. In occasione di un viaggio di studio che si proponeva di affrontare i diversi aspetti della biblioteca digitale in California, organizzato nel 1999 dall'Università Cattolica nell'ambito del master in Gestione e direzione della biblioteca, abbiamo avuto modo di conoscere il progetto LOCKSS (Lots of Copies Keep Stuff Safe).

Iniziato nel 1998 dalle Stanford University Libraries, il LOCKSS Program ha dato luogo a una comunità internazionale di biblioteche che, autorizzate dagli editori, sono in grado di archiviare localmente e preservare proprie copie dei contenuti digitali pubblicati sul web. Ciascuna biblioteca si dota di un LOCKSS box, installando su un proprio server un software open source fornito da Stanford. Mediante un web crawler, il box raccoglie i contenuti digitali, archiviandoli nel formato e nel contesto in cui si trovano sul sito web dell'editore. La leggibilità è garantita

nel tempo attraverso migrazioni dei formati, mentre il confronto con le copie presenti presso le altre biblioteche permette di correggere eventuali differenze. Gli utenti della biblioteca avranno quindi accesso, a seconda dei casi, al contenuto editoriale o alla copia preservata.

I bibliotecari possono definire le unità da preservare, monitorare lo stato della preservazione e controllare l'accesso ai dati salvati.²

Dopo aver analizzato diverse soluzioni, abbiamo deciso di attivare in via sperimentale due iniziative: una di tipo interno, tesa a soddisfare le condizioni sopra enunciate specialmente per le pubblicazioni istituzionali del nostro ateneo; e, nello stesso tempo, l'avvio dalla sperimentazione della tecnologia LOCKSS.

Le tappe della sperimentazione

La nostra biblioteca ha aderito alla LOCKSS Alliance nel 2006. A partire dall'autunno dello stesso anno, abbiamo introdotto nella prassi abituale di tutte le negoziazioni la richiesta che le licenze comprendessero l'autorizzazione all'utilizzo della tecnologia LOCKSS. Il primo accordo in cui è stata inserita questa clausola è quello che abbiamo concluso con l'editore che nel seguito del nostro articolo indicheremo come Editore E.

Ufficializzata l'adesione alla LOCKSS Alliance, abbiamo costituito un gruppo di lavoro in cui fossero presenti competenze, oltre che in materia di acquisizione delle pubblicazioni seriali elettroniche, anche di tipo tecnico-informatico, coinvolgendo il Servizio sistemi informativi della nostra università. Il Servizio sistemi informativi ha curato l'installazione del LOCKSS box UCSC, completandola nel gennaio 2007, e ha collaborato con noi anche in seguito, specie per gli aggiornamenti del software, in dialogo con il sup-

porto tecnico fornito da Stanford. Nel frattempo avevamo individuato gli editori sui quali iniziare la sperimentazione: tre *university presses*, che nel seguito indicheremo come Editori A, B e C, tra cui l'Editore A pubblica solo in formato elettronico, e due editori commerciali, nel seguito indicati come Editori D ed E.

Sempre a gennaio 2007, dall'interfaccia di amministrazione che i sistemisti ci avevano consegnato, potevamo constatare che le prime sei *archival units* relative a periodici dell'Editore C erano state caricate nel nostro box.

Nel giugno 2007 la nostra sperimentazione si è estesa ai periodici dell'Editore D. Nello stesso perio-

do abbiamo contattato l'Editore A perché collaborasse con LOCKSS. Nel novembre 2007, grazie al lavoro svolto nel frattempo dallo staff di LOCKSS, eravamo in grado di aggiungere alle *archival units* del box UCSC anche quelle relative ai periodici di quest'ultimo editore. Sempre nella prima parte dell'anno, ha avuto inizio una serie conti-

Tab. 1

Editore	Host previsto da LOCKSS	N. seriali lavorati	Su un totale abbonamento UCSC per editore	N. <i>archival units</i> per cui il box UCSC è stato settato	Su un totale di <i>units</i> a cui l'UCSC avrebbe diritto di accesso (incl. 2007)	<i>Archival units</i> salvate dal box UCSC	Problemi	Soluzioni
Editore A	Sito editoriale	7	9	37	108	37	-	-
Editore B	Sito aggregatore	1	11	1	99	0	Il nostro accesso è invece sul sito editoriale e sul sito di altro aggregatore.	È necessario: - farsi rilasciare autorizzazione dall'editore e dal nostro fornitore; - ottenere la pubblicazione delle <i>publisher manifest pages</i> ; - ottenere, all'interno dell'Alliance, sei voti per ogni titolo e per ogni piattaforma.
Editore C	Sito editoriale	53	57	127	456	105	1) Cinque <i>archival units</i> ancora in attesa di <i>crawling</i> . 2) Per dodici <i>archival units</i> , "no permission from publisher" (nonostante i diritti di accesso UCSC). 3) Per cinque <i>archival units</i> , "fetch error".	1) Attendere che il <i>crawling</i> venga riavviato. 2) Verificare l'abilitazione dell'indirizzo IP del nostro server LOCKSS. 3) Verifiche tecniche con lo staff di LOCKSS.
Editore D	Sito editoriale	31	112	31	1.008	31	-	-
Editore E	-	-	-	-	-	-	-	L'editore ha dichiarato di non poter rispettare la clausola contrattuale relativa a LOCKSS.

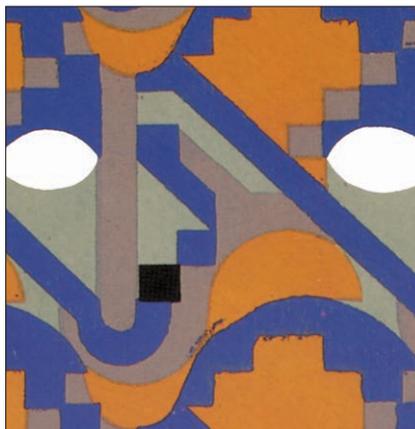
nua di scambi tra la nostra biblioteca e l'Editore E, perché venisse rispettata la clausola contrattuale riferita a LOCKSS. In novembre l'editore ci comunicava definitivamente di non poterla osservare. La tabella 1 sintetizza, anche in termini numerici, le attività svolte, i problemi incontrati e i risultati raggiunti entro novembre 2007.

Gli esiti dopo un anno di attività

I risultati della sperimentazione si sono rivelati per noi buoni sui periodici degli Editori A, C e D. Nel complesso, la tecnologia LOCKSS si è confermata in linea con le strategie seguite dal nostro sistema bibliotecario in tema di preservazione a lungo termine delle pubblicazioni elettroniche e soddisfa le condizioni di cui in premessa.

Nel primo anno di sperimentazione abbiamo così maturato la decisione non solo di rinnovare la *membership* per il 2008, ma anche di supportare la promozione dell'iniziativa a livello nazionale ed europeo. Inoltre, dovendo attivare, a partire dal 2008, abbonamenti a periodici di editori coinvolti nella sperimentazione con esiti positivi, l'Università Cattolica ha stabilito di scegliere la formula *print + online* e di conservare esclusivamente l'edizione elettronica, utilizzando la copia a stampa solo per verificare la conformità dell'edizione online preservata nel LOCKSS box rispetto all'edizione a stampa.

Nel caso di editori già aderenti a LOCKSS e che pubblicano soltanto in formato elettronico, l'Università Cattolica sta ora valutando con interesse l'opportunità di ampliare il numero dei propri abbonamenti, mentre, prima della sperimentazione della tecnologia LOCKSS, tale ampliamento veniva limitato ai casi in cui risultava strettamente necessario.



Per il 2008 abbiamo stabilito di estendere la sperimentazione ai periodici di altri gruppi editoriali, soprattutto *university presses*, e a quelli Open Access. In particolare, per quanto riguarda questi ultimi, che spesso sono fonti informative di notevole rilievo per le attività che la nostra biblioteca è chiamata a supportare, abbiamo deciso di avviare una specifica politica di sviluppo delle collezioni, dedicandovi un'attenzione pari a quella che il nostro sistema bibliotecario ha sempre riservato alle pubblicazioni a stampa accessionate in omaggio o in cambio.

Le criticità di LOCKSS

Una prima criticità che abbiamo incontrato consiste nel fatto che l'individuazione degli *hosts* da cui viene resa possibile la preservazione dipende dagli accordi che le biblioteche dell'Alliance riescono a concludere, di caso in caso, con editori o fornitori. Nessun *host* viene considerato come preferenziale o di riferimento, in linea di principio. Lo staff di LOCKSS predispone l'effettiva preservazione dei periodici elettronici per un determinato *host* nel momento in cui siano stati soddisfatti tre requisiti:

1) che il *content provider* abbia accettato che le biblioteche possano archiviare i dati sui propri server mediante la tecnologia LOCKSS;

2) che il *content provider* abbia pubblicato sul web la *publisher manifest page* per ciascuna *archival unit* (dato volume o annata di un dato titolo);

3) che siano stati ottenuti, all'interno dell'Alliance, almeno sei voti, corrispondenti ad altrettante biblioteche che mettano a disposizione i propri LOCKSS box per le necessarie copie di backup.

La direzione dell'Alliance ha fatto sapere di non volersi ingerire, per motivi comprensibili e corretti, nelle scelte dei fornitori e delle piattaforme, scelte che competono alle singole biblioteche aderenti. D'altra parte, una volta che gli *hosts* per la preservazione di determinati contenuti sono stati definiti, una o più biblioteche dell'Alliance possono vedersi preclusa la possibilità di procedere alla preservazione stessa, pur disponendo dei diritti di accesso a quegli stessi contenuti (su altro/i *host(s)*). Ad esempio, nel caso dei periodici elettronici dell'Editore B che rientrano nei nostri abbonamenti, la biblioteca non ha potuto settare il proprio LOCKSS box per la preservazione delle relative *archival units* perché non dispone dell'accesso a questi periodici sull'*host* a cui LOCKSS fa riferimento, ma ha accesso all'*host* dell'editore.

In presenza di *n* fornitori di cui le biblioteche possano avvalersi per l'accesso a un determinato titolo elettronico, lo staff di LOCKSS dovrebbe, per assurdo, predisporre il settaggio di ciascuna *archival unit* per ognuno degli *n* *hosts*. Una soluzione come questa appare di difficile attuazione e gestione; sarebbe auspicabile che la direzione dell'Alliance stabilisse di considerare sempre l'*host* editoriale come *host* preferenziale e di favorire l'acquisizione dell'autorizzazione all'uso della tecnologia LOCKSS, in primo luogo, da parte dell'editore. Una volta ottenuta l'autorizzazione dell'editore, questa potrà essere fatta

valere presso gli aggregatori che pubblicano i medesimi contenuti editoriali. In tal modo, se anche una biblioteca accede al full text tramite un *host* diverso da quello della maggior parte delle biblioteche dell'Alliance, in presenza dell'autorizzazione dell'editore all'uso della tecnologia LOCKSS è comunque messa nelle condizioni di preservare i contenuti.

Un secondo aspetto di difficoltà da noi rilevato durante il primo anno di sperimentazione è che, confrontato con il complesso di periodici elettronici a cui gli utenti del nostro sistema bibliotecario hanno accesso, il numero di *archival units* preservabili in LOCKSS rappresenta una percentuale minima. Il campione coinvolto nella sperimentazione, compresi i titoli pubblicati dall'Editore E, non raggiunge il 2% del totale dei periodici elettronici del Sistema bibliotecario UCSC. Inoltre, se limitiamo la nostra analisi agli editori che avevano o hanno accettato LOCKSS, sui cinque editori che la nostra biblioteca ha incluso nella sperimentazione, uno non ha eseguito la clausola contrattuale che permetteva l'utilizzo della tecnologia LOCKSS; nel caso di un altro editore, la nostra biblioteca non ha accesso all'*host* utilizzato da LOCKSS per la preservazione. Per gli altri tre editori la sperimentazione ha avuto buoni esiti, ma è stato possibile avviare la preservazione sul LOCKSS box UCSC limitatamente a 92 periodici in abbonamento su 178 (il 51% circa), per complessive 195 *archival units* su 1.572 (il 12,4%).

Occorre poi aggiungere una considerazione riguardante gli editori che aderiscono a LOCKSS. Mancano i grandi editori commerciali (Elsevier, Springer, Wiley), anche se abbiamo notizia di una dichiarazione di interesse verso LOCKSS da parte di Taylor and Francis.

Un'ultima annotazione merita, infine, il problema della preservazione

delle pubblicazioni italiane o di interesse prevalentemente nazionale, nonché, per ciò che riguarda l'Università Cattolica, delle pubblicazioni italiane rilevanti per la ricerca e la cultura cattolica in Italia. A questo proposito, si rende necessaria e urgente la costituzione di una Private LOCKSS Network italiana. L'urgenza è accresciuta dal fatto che il nostro sistema bibliotecario, come altri in Italia, ha dovuto interrompere abbonamenti a diverse decine di seriali a stampa essendosi resa disponibile la sola edizione elettronica distribuita via web.

Conclusioni

Rimaniamo del parere che la tecnologia LOCKSS offra una soluzione molto valida al problema della preservazione dei periodici elettronici, soprattutto per quelli editi da *university presses* e da editori commerciali *library-friends*, per le ragioni che abbiamo già evidenziato. LOCKSS sembra prestarsi però ad ampi sviluppi per le pubblicazioni accessibili gratuitamente sul web. Le questioni che l'utilizzo della tecnologia LOCKSS pone alle biblioteche si possono riassumere così:

- *questioni di ordine tecnico*: sono quelle di più facile soluzione;
- *questioni di ordine contrattuale*: si tratta di acquisire l'autorizzazione dell'editore all'uso della tecnologia LOCKSS; le difficoltà divengono insormontabili se l'editore decide di non sposare questo tipo di soluzione (vedi il caso dell'Editore E);
- *questioni di ordine politico*: affinché lo staff di LOCKSS sviluppi i plug-in necessari alla preservazione delle *archival units* di un periodico, occorre raggiungere un quorum di sei voti (*technical requirement*); per un periodico *free* è richiesto un quorum pari al doppio del *technical requirement*;
- *questioni inerenti allo sviluppo delle collezioni*: è necessario stabili-

re criteri di sviluppo delle collezioni e condividerli con un gruppo critico di altre biblioteche, in collaborazione con le quali sia possibile far fronte più efficacemente alle esigenze prima definite come politiche.

Note

¹ LOCKSS, Lots of Copies Keep Stuff Safe, <<http://www.lockss.org/>>.

² Una breve descrizione di LOCKSS si trova nell'articolo di T. GIORDANO, *Le collezioni non abitano più qui?*, "Biblioteche oggi", 24 (2006), 2, p. 99-100. Per un'introduzione più specifica segnaliamo: K.G. SCHNEIDER, *Lots of librarians can keep stuff safe*, "Library Journal", 132 (2007), 13, p. 30-31. L'articolo di R. FOX, *The double bind of e-journal collections*, "OCLC Systems & Services", 23 (2007), 1, p. 21-29, esamina pro e contro di LOCKSS e Portico (<http://www.portico.org/>). Un'ampia bibliografia è censita sul sito web di LOCKSS, <<http://www.lockss.org/lockss/Publications>>.

Abstract

Like many other libraries, the Library System of Università Cattolica del Sacro Cuore has been confronting the challenges set by electronic publishing. In order to choose their own way in digital preservation, Università Cattolica librarians looked worldwide at the experience of their colleagues, while establishing their own criteria for a conscious transition from print to digital resources.

Here an account is given of the experimentation of LOCKSS technology, started in 2007, involving titles from five publishers. After discussing the issues encountered and summarizing the results achieved during the first year, the authors conclude that LOCKSS technology proves to be a valid and efficient solution for the preservation of library digital collections.