

## La conservazione del digitale: strategie, tecniche, standard ed esperienze per la costruzione di trusted repository di pubblicazioni periodiche \*

Valdo Pasqui

valdo.pasqui@unifi.it

Università di Firenze e AIB Gruppo di studio sulle biblioteche digitali

### Sintesi

*Una parte significativa della nostra eredità culturale generata nell'ambito accademico e scientifico è affidata alle pubblicazioni periodiche in formato elettronico (e-journals) sia di tipo commerciale che ad accesso aperto. Tali pubblicazioni sono soggette alle vulnerabilità tipiche dell'ambiente informatico, in particolare a quelle legate al cambiamento delle tecnologie ed alla sicurezza. È pertanto indispensabile assumere iniziative ed avviare progetti per assicurare alle future generazioni la possibilità di consultare, citare e riutilizzare i contenuti delle riviste elettroniche. La conservazione del digitale (digital preservation) garantisce nel lungo termine l'accesso a tali risorse attraverso l'adozione di opportuni modelli organizzativi e l'applicazione di appropriate politiche, strategie e tecniche. Tale processo deve essere sviluppato con la collaborazione di tutti gli attori coinvolti, le istituzioni accademiche e di ricerca, le biblioteche, gli editori e gli autori, condividendo le necessarie competenze e assicurando il continuo adeguamento attraverso una costante azione di verifica e valutazione. Questo contributo intende fornire un quadro di riferimento sulla conservazione a lungo termine delle riviste digitali. Dopo aver richiamato le principali caratteristiche degli archivi affidabili (trusted repository) ed i capisaldi della digital preservation, si evidenziano alcune specificità delle riviste elettroniche. Poi si analizzano le soluzioni, le strategie e le tecniche adottate da quattro programmi internazionali che rappresentano altrettanti significativi e diversi approcci per la conservazione delle riviste elettroniche. Infine vengono riportate le principali risultanze pratiche del rapporto prodotto nel 2006 dalla Association of Research Libraries (ARL) e dal Council on Library and Information Resources (CLIR) e si indicano alcuni punti fermi formulando l'ipotesi di una rete nazionale di archivi affidabili.*

### 1. Importanza della *digital preservation*

Nel corso dell'ultima decade al fine di ridurre i costi di stampa, distribuzione e gestione dei periodici si è verificato uno spostamento sempre più significativo verso l'editoria elettronica sia da parte degli editori che delle biblioteche, in particolare quelle che operano negli ambiti accademici e di ricerca. Una misura indicativa di questa realtà è data dal confronto tra il "Directory of Scholarly Electronic Journals and Academic Discussion Lists" [Mogge], pubblicato dalla Association of Research Libraries' (ARL), che a fine 2000 aveva censito 3.915 riviste, e il numero dei titoli dell'archivio delle riviste elettroniche e delle newsletter mantenuto da NewJour [NEWJOUR] che al 31 Ottobre 2007 riporta ben 20.377 titoli. In un documento pubblicato dalla Scholarly Publishing and Academic Resources Coalition (SPARC) nel 2006 [Crow] si afferma che circa il 60% delle 20.000 riviste soggette a peer-review sono ormai disponibili in formato elettronico, mentre [Chandra] riporta che nelle biblioteche universitarie appartenenti ad ARL dal 2002 al 2006 il numero dei titoli sottoscritti solo in formato cartaceo si è ridotto dal 64% al 30% con un incremento

---

\* Questo articolo è il pre-print dell'intervento tenuto al "Convegno Collezioni digitali di periodici in Italia e in Europa: standard, applicazioni, valutazioni, prospettive" tenutosi ad Arco (Trento) il 15-16 Novembre 2007 ed i cui atti sono in corso di stampa.

delle sottoscrizioni di tipo “*e-only*” dal 5% al 37%. Infine, un recente rapporto ARL afferma che “*Except for top-tier, broad circulation titles—which sometimes are used more like magazines—surviving printed editions may become mere add-ons available via print on demand*” e aggiunge “*The role of the printed journal in the institutional marketplace faces a steep decline in the coming 5 to 10 years. Print journals will exist mainly to address specialized needs, users, or business opportunities*” [Johnson&Luther].

A questo scenario va aggiunto il numero crescente di repository istituzionali ad accesso aperto che contengono in formato elettronico i *pre-prints* e *post-prints* di articoli di riviste oltre al materiale didattico, ai rapporti ed ai documenti prodotti nell’attività di ricerca.

La sempre più accentuata dipendenza dai formati digitali obbliga a valutarne i principali fattori di rischio: la fragilità dei documenti elettronici e i vincoli derivanti dalle licenze d’uso per gli *e-journals*. Il primo è determinato dalla forte dipendenza che le pubblicazioni elettroniche e gli oggetti digitali hanno rispetto al rapido cambiamento, alla obsolescenza e alle vulnerabilità delle piattaforme tecnologiche hardware e software che li ospitano. Il secondo è legato alla forte concentrazione delle pubblicazioni elettroniche nelle mani di pochi editori ed aggregatori i quali forniscono l’accesso remoto agli articoli applicando modelli contrattuali che spesso non garantiscono con certezza il diritto di mantenere l’accesso perpetuo alle riviste sottoscritte, sia relativamente ai numeri pregressi in caso di cancellazione dell’abbonamento sia per situazioni “critiche” dovute all’editore stesso. In particolare gli utenti rischiano di perdere la possibilità di usare queste pubblicazioni quando l’editore per ragioni tecniche non riesce a garantire l’accesso a tutte o parte delle proprie riviste elettroniche per periodi di tempo più o meno lunghi, oppure se dismette alcuni titoli o il proprio archivio o se cessa l’attività.

Il 13 Settembre 2005 un gruppo di esperti appartenenti a varie università, biblioteche e centri di ricerca americani, riunito a New York presso The Andrew W.Mellon Foundation, ha discusso i problemi riguardanti la conservazione delle riviste elettroniche producendo un documento intitolato “Urgent Action Needed to Preserve Scholarly Electronic Journals” [URGENT-ACTION]. Si tratta di una sorta di manifesto che sottolinea l’urgenza e la necessità della conservazione delle riviste elettroniche attraverso la creazione di :

*“...trusted archives in which the published scholarly record in electronic form can persist outside of the exclusive control of publishers, and in the control of entities that value long-term persistence”*

Il documento individua quattro linee di azione da attivare al fine di perseguire tale obiettivo:

- a) considerare la *digital preservation* delle riviste elettroniche come una forma di assicurazione contro il rischio della loro perdita permanente e comunque per sopperire ad eventuali interruzioni prolungate dell’accesso dovute a problemi degli editori;
- b) fare in modo che gli archivi destinati alla conservazione offrano un insieme minimale di servizi ben definiti (ricezione e memorizzazione dei file, accesso ed utilizzo) accompagnati da un costante e trasparente processo di controllo e verifica;
- c) promuovere gli investimenti da parte delle biblioteche per la creazione e gestione, anche in forme cooperative, di archivi di questo tipo che costituiscano per le istituzioni partecipanti un sorta di assicurazione mutua e collettiva;
- d) fare pressioni sugli editori, anche attraverso le maggiori associazioni (ALA, ARL, IFLA), affinché si impegnino a garantire il deposito ai fini dell’archiviazione e perché accettino di inserire nei contratti che regolano le licenze dei periodici elettronici una clausola che li obbliga a trasmettere tutti i file per la conservazione in un archivio qualificato.

Queste indicazioni coinvolgono livelli diversi, la prima e l'ultima quello politico e finanziario, le altre due quello organizzativo e strategico, e sono rivolte alla comunità accademica e di ricerca e alle biblioteche affinché attivino programmi per la conservazione a lungo termine delle pubblicazioni digitali per assicurare che questi contenuti possano sopravvivere nel tempo e che sia garantita la possibilità di usarli e citarli.

Prima di passare ad esaminare come alcuni servizi hanno dato risposta a questo appello è opportuno richiamare i punti chiave della conservazione del digitale nel contesto delle riviste elettroniche.

## 2. Archivi affidabili e *digital preservation*

In base alla definizione elaborata dal JISC la *digital preservation* è:

*“the series of actions and interventions required to ensure continued and reliable access to authentic digital objects for as long as they are deemed to be of value”* [JISC].

Elemento centrale di questo processo sono i *“trusted digital repository”*:

*“A trusted digital repository is one whose mission is to provide reliable, long-term access to managed digital resources to its designated community, now and in the future. Trusted digital repositories may take different forms: some institutions may choose to build local repositories while others may choose to manage the logical and intellectual aspects of a repository while contracting with a third-party provider for its storage and maintenance”* [RLG-OCLC]

Nel rapporto pubblicato da RLG-NARA nel 2005 e aggiornato nel Febbraio 2007 dal Center for Research Libraries (CRL) [TRAC] si sottolinea che la corretta formulazione inglese dovrebbe essere *“trustworthy”*, anche se nell'uso comune ormai si è consolidata l'espressione *“trusted”*. Ci sembra appropriato tradurre questo termine con l'aggettivo *“affidabile”* e opportuno sottolineare le principali proprietà cui deve soddisfare un archivio per essere definito affidabile:

Proprietà	Descrizione	Tecniche
Autenticità ( <i>Authenticity</i> )	Riguarda la certezza che una risorsa digitale appartenga a chi dichiara di esserne l'autore (creatore) e corrisponda a quanto dichiara di contenere. Implica anche la proprietà di <i>“non ripudio”</i> cioè la certezza che l'autore non possa negare di essere il creatore dell'oggetto.	La firma digitale basata su l'infrastruttura a chiave pubblica (PKI) e il digital watermarking sono le tecniche più usate per assicurare l'autenticità degli oggetti digitali.
Integrità ( <i>Integrity</i> ).	Concerne la inalterabilità e la stabilità degli oggetti digitali. Garantire l'integrità dei dati significa prevenire e controllare le perdite e i cambiamenti dovuti a cause accidentali o dolose.	La generazione di sequenze di tipo bit checksum (e.g. MD5) è la tecnica adottata per verificare se a distanza di tempo dall'archiviazione qualche modifica abbia alterato la sequenza di bit che formano un oggetto digitale.
Affidabilità ( <i>Reliability</i> ) e Disponibilità ( <i>Availability</i> )	Questi due concetti riguardano la consistenza e la completezza dei dati. Il primo si riferisce al comportamento dei componenti hardware e software dell'archivio che deve essere conforme alle loro specifiche garantendo una percentuale di errori ed anomalie minimale. Il secondo riguarda il regolare funzionamento dei sistemi e delle le applicazioni che devono produrre costantemente i medesimi risultati al	Salvataggi, antivirus, firewall, patch dei sistemi operativi, aggiornamenti dei software applicativi, componenti hardware ridondanti e fault tolerant sono tecniche e soluzioni adottate per massimizzare queste caratteristiche. Le percentuali d'errore vengono di solito calcolate valutando i servizi ritenuti critici rispetto al tempo totale di funzionamento o di erogazione.

	fine di assicurare la corretta fruizione degli oggetti digitali archiviati e dei dati ad essi associati.	
<i>Accesso e utilizzo continuato</i>	Questi concetti si riferiscono alla esigenza di garantire l'uso delle risorse digitali nel tempo ( <i>ongoing usability</i> ) assicurandone l'accesso perpetuo, l'usabilità e la citabilità nel lungo termine. Rientra in questo ambito anche l'accesso " <i>post cancellation</i> ", cioè dopo la scadenza della licenza d'uso, alle riviste elettroniche.	Per garantire queste proprietà i sistemi per la conservazione applicano opportune strategie per il mantenimento e la trasformazione dei formati degli oggetti digitali e si avvalgono di identificatori univoci e persistenti.

**Tabella 1 – Proprietà di un archivio “affidabile”**

La definizione di archivio affidabile distingue chiaramente le finalità dal modello organizzativo adottato per perseguirle e lascia aperta la scelta tra un approccio completamente autonomo e soluzioni che demandano la conservazione ad un servizio esterno. Infatti i termini “*digital repository*” o “*digital archive*” non devono essere intesi solo in senso tecnico-informatico, ma sono riferiti ad organizzazioni che si assumono precisi impegni riguardo alla conservazione nel lungo termine delle risorse digitali: “*an organization that intends to preserve information for access and use by a Designated Community*” [OAIS].

La conservazione del digitale si fonda quindi sulla creazione e sulla gestione di archivi digitali “affidabili” ed è un processo complesso e articolato che richiede: i) una puntuale definizione degli obiettivi e un’attenta pianificazione; ii) un’organizzazione ben strutturata; iii) appropriate politiche, strategie e tecnologie e iv) un monitoraggio continuo che assicuri la corretta applicazione di metodi e procedure e l’adeguamento agli standard correnti.

Le prime due componenti riguardano il livello politico-organizzativo e sono valutabili secondo criteri di sostenibilità finanziaria, efficienza e vitalità dell’archivio. Un ruolo determinante è svolto dalla comunità di riferimento che affida all’archivio il mandato di operare. Denominata “*Designated Community*” [OAIS], comprende i potenziali consumatori delle risorse digitali, ai quali devono essere assicurati in modo persistente nel tempo l’usabilità e la comprensibilità delle risorse, e gli stessi editori che detengono i diritti sulle pubblicazioni e che possono affidare la conservazione nel lungo periodo a consorzi di biblioteche o a servizi terza parte.

Le politiche e le strategie devono fondarsi su appropriati modelli e avvalersi degli standard tecnologici. Il “Reference Model for an Open Archival Information Systems” (OAIS) ha formalizzato: a) il modello funzionale che specifica i processi (*ingest, archival storage, preservation planning & strategies, data management, access management*) e b) il modello informativo che individua i metadati per la conservazione [OAIS].

La componente tecnologica include gli apparati hardware e le applicazioni software e comprende l’attuazione delle politiche necessarie a garantire la sicurezza delle infrastrutture informatiche, telematiche e logistico-ambientali del sito che ospita l’archivio.

Tutte le procedure adottate devono essere documentate in modo esaustivo e sottoposte ad un costante processo di verifica (*audit*) al fine di poter certificare l’affidabilità dell’archivio e per garantire alla comunità di riferimento l’allineamento rispetto all’evolversi delle tecnologie e degli standard. I criteri di certificazione elaborati dal Council on Library and Information Resources [TRAC] e l’analogo “catalogo dei criteri” del progetto tedesco *nestor* [NESTOR] definiscono una griglia di elementi per verificare e valutare l’affidabilità di un archivio ai fini della sua certificazione come “*trusted repository*”, mentre la procedura elaborata dal The Digital Curation Centre (DCC) e dal Digital Preservation Europe (DPE) [DRAMBORA] prevede l’individuazione e la valutazione dei rischi ispirandosi ai modelli adottati per valutare la sicurezza dei sistemi informatici.

Gli aspetti organizzativi e tecnici per la conservazione dei periodici elettronici si inquadrano in questo schema di riferimento non senza ricordare alcune caratteristiche peculiari di queste pubblicazioni che richiedono un trattamento specifico.

In primo luogo i contenuti delle riviste elettroniche non sono solo gli articoli, ma comprendono anche altre componenti molto diversificate: gli editoriali, le rubriche dedicate alle recensioni, alla presentazione di prodotti e alla segnalazione di congressi ed eventi, le sezioni che ospitano la corrispondenza con i lettori, quelle dedicate alle sintesi di avvenimenti ed eventi e le inserzioni pubblicitarie. Conservare solo i singoli articoli, per esempio in formato PDF, o effettuare l'archiviazione completa del sito della rivista elettronica sono approcci molto diversi che implicano l'adozione di appropriate strategie di conservazione (vedi par.5). Alcune soluzioni (e.g. PORTICO) hanno optato sia per il mantenimento degli oggetti digitali originali sia per la loro trasformazione in un formato normalizzato.

Un'altra decisione rilevante riguarda se conservare il *look and feel* originario della rivista, al quale sono abituati i lettori, o rinunciarvi limitandosi a cogliere solo i contenuti intellettuali. Queste scelte devono essere valutate in funzione della complessità e dei costi del processo di conservazione. Infine una peculiarità della conservazione delle riviste elettroniche è il mantenimento delle relazioni strutturali tra le componenti che formano il contenuto informativo e delle citazioni bibliografiche che spesso sono riferimenti (*link*) ad altre risorse elettroniche. L'uso di identificatori univoci e persistenti ed i relativi servizi di risoluzione sono indispensabili per assicurare un livello adeguato di accesso, usabilità e navigazione dei contenuti archiviati.

Il mantenimento della veste grafica e della struttura di una rivista elettronica sono anche strettamente collegati ai requisiti di accesso da parte della comunità di riferimento esprimibili attraverso la domanda: quando, a chi e come vengono resi disponibili i contenuti archiviati ai fini della preservazione a lungo termine?

La Digital Library Federation nei criteri minimi per la conservazione delle riviste accademiche [MINIMUM-CRITERIA] ha risposto indicando che: *“Although repositories will need to support access at some level, those services should not replace the normal operating services through which digital scholarly publications are typically made accessible to end users.”*

Esistono tre modalità diverse per regolare l'accesso ai contenuti conservati che vengono rappresentate dalla diversa tonalità di colore attribuita all'archivio: i) il nero (*dark archive*) [Pearce-Moses] indica che non esiste alcun accesso corrente alle risorse se non da parte dello staff gestionale; ii) il chiaro (*light archive*) indica la possibilità di accedere al contenuto senza alcuna limitazione; iii) l'opaco (*dim archive*) è la soluzione intermedia. Verso quest'ultimo approccio si sono orientati gli estensori della “Urgent Action” che hanno previsto l'accesso solo quando *“(a) the publisher goes out of business or is otherwise unable to provide consistent access; or (b) the content is no longer protected by copyright”*. In questo caso l'archivio implementa alcuni servizi per l'accesso ai contenuti conservati che però vengono resi disponibili alla comunità solo al verificarsi di determinati eventi (*trigger event*) precedentemente concordati con gli editori. Lo studio effettuato dal Council on Library and Information Resources [METES] che tratteremo più avanti ha individuato gli eventi che possono “attivare” l'accesso all'archivio: l'editore cessa di erogare servizi; l'editore non fornisce più i vecchi numeri della rivista; è terminato il copyright sulla rivista; la rivista non viene più pubblicata; l'editore o il servizio di distribuzione hanno un problema temporaneo o grave (“catastrofico”) che impedisce l'accesso alle riviste.

Occorre poi osservare che lo scenario delle riviste elettroniche è comunque in continua evoluzione grazie al diffondersi di tecniche basate sulla generazione dinamica, il riuso e la combinazione dei contenuti (*data and content mashup*, RSS e Yahoo! Pipes, *overlay journals*, *object reuse & exchange*). In un recente articolo Clifford Lynch [Lynch] ha affrontato la tematica della rapida trasformazione della forma degli articoli scientifici mettendo in evidenza alcuni aspetti: a) l'esistenza sempre più frequente di “materiali supplementari” (insiemi di dati, programmi), spesso localizzati esternamente, che accompagnano gli articoli e che pongono questioni tecniche (utilizzo di appropriati schemi di identificazione e di sistemi di risoluzione per localizzare questi oggetti) e

implicano la necessità di estendere i processi di *peer review* alla verifica della validità dei dati di tali pubblicazioni; b) l'uso di software aggiuntivi che elaborano e visualizzano questi dati; c) la progressiva diffusione di tecniche di analisi computazionale, confronto, correlazione, analisi delle anomalie denominate “*data mining*” e “*text mining*” che operano su ampi corpora di letteratura scientifica. Anche grazie alle possibilità offerte da tali approcci alcuni ambiti accademici e scientifici sono sempre più orientati verso riviste “*e-only*” perchè permettono di realizzare pubblicazioni molto ricche e articolate. Ma tali evoluzioni hanno un impatto significativo sui processi di conservazione. L'articolo non è più riconducibile ad un testo sequenziale (es. in PDF) ma diventa un oggetto composito e complesso. Questa evoluzione pone nuove sfide alla *digital preservation* perchè la conservazione di tutte le sue componenti, in particolare di quelle generate dinamicamente, e delle loro relazioni è determinante ai fini di garantire la consistenza e la completezza del contenuto informativo.

### 3. Servizi per la digital preservation delle riviste elettroniche

Nel 2006 la Association of Research Libraries (ARL) e il Council on Library and Information Resources (CLIR) hanno pubblicato un rapporto [METES], elaborato dal dipartimento Library Research and Assessment Services della Cornell University, che analizza lo stato dell'arte delle principali iniziative di archiviazione degli e-journals. Lo studio ha censito dodici iniziative che gestiscono archivi aventi per finalità la conservazione nel lungo termine delle riviste elettroniche. Alcuni dei programmi sono stati avviati già da molti anni come quelli del Los Alamos National Laboratory Research Library (LANL-RL) iniziato nel 1994 e della National Library of Australia (NLA) PANDORA iniziato nel 1996, altri sono più recenti come Portico, LOCKSS Alliance e CLOCKSS. In gran parte si tratta di progetti finanziati con fondi governativi, due sono consorzi (OhioLINK Electronic Journal Center e Ontario Scholars Portal) e quattro prevedono anche il pagamento da parte delle istituzioni partecipanti (Portico, OCLC Electronic Collections Online, LOCKSS Alliance and CLOCKSS). Sono iniziative europee e-Depot della Koninklijke Bibliotheek in Olanda e il Kooperativer Aufbau eines Langzeitarchivs Digitaler Informationen (kopal/DDB) in Germania.

Dall'analisi dei modelli organizzativi e delle soluzioni tecniche adottate emergono quattro approcci diversi che nel seguito descriviamo brevemente attraverso le caratteristiche essenziali del servizio più rappresentativo per ciascuna categoria:

- il dark-archive gestito da una organizzazione terza parte - Portico
- l'archivio decentralizzato gestito da biblioteche cooperanti - LOCKSS Alliance
- l'approccio istituzionale per il deposito legale e la conservazione - e-Depot
- l'archivio di conservazione ad accesso libero - Pub Med Central

Portico (URL: <http://www.portico.org/>), promosso e finanziato da JSTOR, da The Andrew W. Mellon Foundation, dalla Ithaka Harbors Inc. e dalla Library of Congress, è stato avviato nel 2005 e reso operativo nel 2007 e si propone come servizio terza-parte per le biblioteche e per gli editori attraverso la sottoscrizione di licenze. Il servizio è finanziato mediante le quote di contribuzione versate dai partecipanti: le biblioteche contribuiscono versando in funzione della loro spesa annua per l'acquisto di materiale librario; gli editori versano un contributo annuale, calcolato in funzione dei loro ricavi. Portico è un *dark archive* che consente l'accesso alle biblioteche sottoscrittrici se si verifica uno dei *trigger event* previsti oppure se l'editore ha scelto Portico come servizio per garantire l'accesso di tipo post-cancellazione o perpetuo alle proprie riviste, infatti non è prevista la rimozione dei contenuti forniti. Questa caratteristica viene esplicitamente dichiarata nel messaggio che compare nella pagina di accesso ad una rivista: “*Full access to the Portico archive is intended for verification and testing purposes and is limited to library staff and publishers. Distributing materials from the archive to library patrons or otherwise access for any purpose other than*

*verification and testing is not permitted.*” È interessante notare che i trigger event attivano l’accesso di tipo campus wide per tutte le biblioteche sottoscrittrici del servizio, indipendentemente che queste abbiano o no sottoscritto in precedenza la licenza con l’editore. È interessante notare che i trigger event attivano l’accesso di tipo campus wide per tutte le biblioteche sottoscrittrici del servizio, indipendentemente dal fatto che abbiano o no sottoscritto in precedenza la licenza con l’editore. La strategia di conservazione di Portico prevede di acquisire dagli editori i file sorgenti e li trasforma in un formato normalizzato per l’archiviazione. L’attenzione è rivolta a mantenere e preservare il contenuto intellettuale (testi, immagini, tabelle e relazioni interne alla rivista) senza catturare e conservare l’aspetto grafico ed altre funzionalità fornite dal sito on-line dell’editore. I file sorgenti di solito sono forniti in formato SGML o XML e vengono trasformati in file XML conformi alla Journal Archiving and Interchange DTD definita dalla National Library of Medicine. Poi vengono estratti e aggiunti i metadati necessari per la conservazione secondo lo standard PREMIS. Infine per ogni articolo tutti i componenti necessari alla riproduzione vengono impacchettati in un’unità informativa di archiviazione (AIP di OAIS) che è una variante dello schema XML METS (Metadata Encoding & Transmission Standard). Con questo approccio Portico cerca di limitare i costi, infatti il contenuto viene riprodotto e visualizzato interpretando il formato normalizzato, basandosi sulle caratteristiche standard dei browser web attuali.

"Lots of Copies Keep Stuff Safe" (LOCKSS) (URL: <http://www.lockss.org/> ) è un programma lanciato nel 1999 dalla Stanford University Library il cui logo, una tartaruga, animale notoriamente molto longevo, indica in modo significativo la sua missione. Il sistema è entrato in produzione nel 2004. La maggior parte dei contenuti sono accessibili solo ai membri della LOCKSS Alliance. Quest’ultima è un’organizzazione aperta, costituita e governata dalle biblioteche partecipanti che contribuiscono finanziariamente allo sviluppo del software e del programma ed alla promozione di reti private basate sul software LOCKSS. Le biblioteche membro possono accedere ai contenuti degli editori con i quali hanno in precedenza stipulato una licenza o l’acquisto. In Italia fanno parte di LOCKSS Alliance (a fine 2007) l’Istituto di Scienza e Tecnologie dell’Informazione (ISTI) del CNR e la Biblioteca d’Ateneo dell’Università Cattolica del Sacro Cuore a Milano. Il modello è completamente distribuito e utilizza una serie di computer dedicati, posti in rete ed equipaggiati con un apposito software che li trasforma in un *appliance*, ovvero in una macchina specializzata a svolgere un solo e ben determinato compito, per mezzo di un software *open source* di facile installazione. Questa soluzione, denominata “LOCKSS Box”, è stata specificatamente concepita per fornire alle biblioteche un “*easy and inexpensive way to collect, store, preserve, and provide access to their own, local copy of authorized content*”. Per aumentare il livello di sicurezza ciascuna appliance opera tramite un boot CD che contiene il sistema operativo (derivato da OpenBSD) e tutti i programmi software ed usa una flash-memory USB (od un floppy) protetta in scrittura per contenere la configurazione, mentre i dischi fissi sono utilizzati solo per ospitare i dati archiviati. L’acquisizione dei contenuti delle riviste da archiviare prevede l’accesso ai siti degli editori secondo un approccio di tipo web crawling come i motori di ricerca internet. La LOCKSS box dispone di un motore di ricerca e di una serie di moduli (“Plugin”) specializzati, scritti dagli editori e dagli amministratori dei sistemi appartenenti alla comunità e configurabili con le informazioni necessarie per catturare il contenuto delle riviste di un editore (sito dell’editore, frequenza di pubblicazione e URL per iniziare la ricerca/cattura di una rivista, struttura e contenuto delle riviste, regole relative all’aggiornamento dei contenuti per guidare il sistema di cache). Naturalmente ogni editore deve fornire le opportune autorizzazioni per consentire ad un LOCKSS box di catturare e conservare le proprie riviste, attraverso la pubblicazione sul proprio sito di una pagina web detta “LOCKSS publisher manifest” che permette al motore di ricerca di recuperare alcuni metadati ed il punto di partenza (URL) da cui iniziare la ricerca per la cattura degli articoli. Esiste una pagina di questo tipo per ogni “Archival unit” abilitata alla conservazione che di solito corrisponde ad un volume per le riviste soggette a sottoscrizione o all’intera rivista se si tratta di una pubblicazione ad

accesso aperto. Di norma i permessi sono contrattati con gli editori dai membri della LOCKSS Alliance e poi sono estesi a tutte le istituzioni aderenti autorizzate.

L'accesso agli articoli avviene attraverso l'uso di un normale browser web e sfrutta il principio di funzionamento dei ben noti proxy server. Una biblioteca che ha attivato un LOCKSS appliance configura il proprio proxy server in modo da intercettare le richieste destinate al sito dell'editore e se tale sito non risponde il proxy server redirige il browser al sistema LOCKSS che restituisce la copia conservata. Pertanto, fatta eccezione per gli oggetti generati dinamicamente, l'utente vede lo stesso contenuto che sarebbe fornito dal sito dell'editore. L'accesso è perpetuo perché i contenuti archiviati non sono rimossi dalla LOCKSS box e quindi questa soluzione fornisce anche l'accesso di tipo post cancellazione.

Nel 2006 è stato avviato CLOCKSS (Controlled LOCKSS) un'iniziativa *non profit* che attualmente è formata da comunità di sei biblioteche americane ed una inglese, ciascuna delle quali ospita due LOCKSS box che provvedono all'archiviazione di contenuti di editori accademici per i quali le biblioteche possono avere o no la sottoscrizione, controllandone e mantenendone continuamente l'integrità. Questi sistemi formano un unico *dark archive* distribuito in rete e quando il board che governa il progetto accerta il verificarsi di un *trigger event* i relativi contenuti sono trasferiti su opportune piattaforme e posti ad accesso libero per chiunque. Tale modello è replicabile anche in altri gruppi "privati" di biblioteche.

La Koninklijke Bibliotheek (KB) è la Biblioteca Nazionale olandese responsabile del deposito legale e della conservazione delle pubblicazioni in Olanda. Fin dal 1996 ha operato attraverso specifici accordi con i maggiori editori olandesi (Elsevier Science e Kluwer Academic) per attivare programmi di conservazione a lungo termine degli e-journal pubblicati in Olanda. Dopo aver partecipato con altre biblioteche nazionali europee al progetto NEDLIB (1998-2000) nel 1999 ha affidato ad IBM la realizzazione di un sistema di deposito elettronico. Ha così origine il Digital Information and Archiving System (DIAS) su cui si basa e-Depot, operativo dal 2003, che può ricevere dagli editori pubblicazioni sia on-line che off-line (p.e. su CD-ROM o DVD). Per il momento e-Depot (URL: <http://www.kb.nl/dnp/e-depot/e-depot-en.html>) prevede solo il deposito di pubblicazioni elettroniche in formato PDF, mentre per quelle off-line viene effettuata un'installazione locale la cui immagine completa, comprendente oltre al contenuto anche il software applicativo e il sistema operativo, viene archiviata in DIAS in modo da poterla successivamente ripristinare su una postazione di lavoro per consentirne l'uso. e-Depot opera come "dark-archive" concepito in primo luogo per assicurare agli editori un sistema sicuro per la conservazione delle loro pubblicazioni: "*the KB has decided to make its facilities for long-term preservation available to international scientific publishers*". Infatti non è prevista alcuna forma di accesso al materiale depositato (salvo gli oltre cento titoli ad accesso aperto di BioMed Central) e il *document delivery* viene effettuato solo se è stato previsto dagli accordi con l'editore, è invece possibile chiedere la stampa di porzioni limitate di un articolo per uso personale. Le pubblicazioni elettroniche depositate possono essere consultate solo da personale autorizzato (membri dello staff o persone appositamente autorizzate dalla KB) da postazioni che operano al di fuori della rete pubblica della Biblioteca Nazionale. Tuttavia in caso di una prolungata indisponibilità dei servizi dell'editore o nel caso in cui l'editore smettesse di rendere accessibile alcune riviste, la KB potrebbe renderle disponibili sia localmente che con accesso remoto.

DIAS è utilizzato anche da KOPAL i cui partner sono Die Deutsche Bibliothek (la Biblioteca Nazionale della Germania) e SUB Göttingen (la biblioteca dell'Università e dello Stato della Bassa Sassonia) che è un programma attivato nel 2004 per la conservazione di risorse elettroniche incluse anche le riviste sulla base di accordi con alcuni editori.

Il National Center for Biotechnology Information (NCBI) della National Library of Medicine (NLM) ha sviluppato e gestisce PubMed Central (PMC) allo scopo di attuare le politiche di digital preservation del National Institutes of Health (NIH). PubMed Central (URL:

<http://www.pubmedcentral.nih.gov/>) è stato avviato nel 2000 e archivia le pubblicazioni elettroniche concernenti l'area biomedica. L'accesso libero e gratuito ai contenuti è la sua caratteristica distintiva, infatti la politica promossa da PubMed Central nei confronti degli editori delle riviste prevede di: a) ottenere il contenuto completo di ogni numero a poca distanza di tempo dalla pubblicazione, anche nei casi in cui non viene rilasciato per la libera consultazione; b) renderlo visibile il prima possibile, entro un anno gli articoli ed entro tre anni dalla data di pubblicazione per tutti gli altri contenuti.

La NLM ha anche lanciato PMC International (PMCI) una iniziativa internazionale che coinvolge organizzazioni di altri paesi interessate alla conservazione delle pubblicazioni digitali di ambito biomedico e che ha come obiettivo la creazione di una rete di archivi digitali in grado di condividere tra loro tutto o parte del proprio contenuto. È previsto che ciascuno di questi archivi sia gestito da un'organizzazione nazionale o internazionale di tipo governativo o non profit e che gli editori sottoscrivano uno specifico accordo con la NLM per consentire l'accesso libero ai contenuti depositati e la loro distribuzione agli altri archivi che compongono PMCI. A loro volta gli archivi che fanno parte di PMCI devono sottoscrivere con la NLM un accordo formale per assicurare di mantenere l'integrità dei contenuti, di proteggere i diritti degli editori sul materiale depositato e di garantire agli utenti sempre l'accesso gratuito.

La Tabella 2 riassume e mette a confronto alcune delle caratteristiche più significative delle quattro approcci di conservazione delle riviste digitali.

	<b>Portico</b>	<b>LOCKSS Alliance</b>	<b>e-Depot</b>	<b>Pub Med Central</b>
<b>Tipologia</b>	dark-archive gestito da una organizzazione terza parte	archivio distribuito tra biblioteche cooperanti	deposito legale e conservazione	archivio di conservazione ad accesso libero corrente
<b>Acquisizione dei contenuti</b>	forniti dagli editori, il contenuto intellettuale (testi, immagini, tabelle e relazioni interne alla rivista) viene prelevato senza conservare il "look-and-feel"	prelevati dai siti degli editori con tecniche di web crawling "guidato" dalle archival unit catturando tutto ciò che viene esposto	forniti dagli editori, pubblicazioni on-line (PDF) e off-line (CD-ROM,DVD)	gli editori forniscono il contenuto completo di ogni numero entro poco tempo dalla data di pubblicazione
<b>Formati</b>	i file sorgenti forniti in formato SGML o XML sono normalizzati in file XML conformi alla Journal Archiving and Interchange (JAI) DTD	il contenuto è catturato come viene esposto ("as is") dai siti degli editori	sorgenti in formato PDF e immagini complete di CD-ROM e DVD	il National Center for Biotechnology Information della NML ha formalizzato la Journal Archiving and Interchange (JAI) DTD
<b>Strategie di conservazione</b>	normalizzazione del formato (JAI DTD), metadati PREMIS, archival information package conforme a METS (Metadata Encoding & Transmission Standard )	migrazione dei formati, rete di "appliance" (LOCKSS box) cooperanti tra loro per verificare l'integrità e la completezza delle risorse	migrazione dei formati, emulazione con UVC (Universal Virtual Computer)	normalizzazione dei formati (JAI DTD) e migrazione
<b>Modalità di accesso</b>	"dark archive": consente l'accesso alle biblioteche sottoscrittrici se si verifica uno dei "trigger event" previsti e se l'editore ha scelto Portico come servizio per l'accesso di tipo post-cancellazione o perpetuo	la biblioteca che ha un LOCKSS appliance configura il proprio proxy server che intercetta le richieste destinate al sito dell'editore e se questo non risponde redirige il browser al sistema LOCKSS che restituisce la copia conservata	"dark-archive", accesso al materiale depositato solo al personale autorizzato e da postazioni della KB, DD solo se previsto dagli accordi con l'editore	"light-archive": accesso libero e gratuito, ma alcuni articoli vengono resi disponibili dopo X mesi in base agli accordi con gli editori

<b>Modalità di adesione al servizio</b>	contribuzione annua: - biblioteche in funzione della loro spesa annua per l'acquisto di materiale librario (5 anni) - editori in funzione dei ricavi derivanti dalle licenze e dalla pubblicità per le riviste (3 anni)	la biblioteca: - installa il sw open-source LOCKSS e trasforma un PC in un appliance; - stipula gli accordi con gli editori; - condivide le risorse nell'ambito di LOCKSS Alliance o di gruppi "privati"	servizio di deposito legale della Biblioteca Nazionale Olandese, accordi con gli editori	accordi stipulati con gli editori per le pubblicazioni dell'area biomedica
<b>Piattaforma tecnologica</b>	sistema proprietario conforme a OAIS, orientato a garantire l'usabilità dei contenuti tramite browser web	rete distribuita di appliance basati su PC standard, sw open source (include s.o. derivato da OpenBSD)	prodotto commerciale Digital Information and Archiving System (DIAS) sviluppato dalla IBM	sistema proprietario
<b>Dati rilevati al 31/10/2007</b>	370 biblioteche, 41 editori oltre 6.000 titoli circa 500.000 articoli	184 biblioteche (2 italiane) 170 editori	12 editori oltre 5.000 titoli, circa 10 milioni di articoli	358 titoli di riviste delle quali 292 con rilascio immediato in PMC dalla data di pubblicazione e di queste 214 ad accesso aperto

**Tabella 2 “Soluzioni per la conservazioni di riviste digitali”**

#### **4. Modelli architetturali e organizzativi: verso un'infrastruttura a rete**

Tre delle soluzioni esaminate, Portico, e-Depot e PubMed Central, seguono il modello centralizzato e soltanto LOCKSS è stato concepito secondo un approccio totalmente distribuito. In particolare e-Depot si basa su DIAS, un prodotto sviluppato e commercializzato da IBM, che utilizza costosi e sofisticati componenti tra i quali Tivoli Storage Manager, il sottosistema di archiviazione a più livelli che IBM propone per i data center di grosse dimensioni. Al contrario LOCKSS si basa su un insieme di sistemi tra loro cooperanti in rete, ciascuno dei quali consiste in un PC che viene trasformato in un *appliance*, ovvero in una macchina specializzata a svolgere un solo e ben determinato compito, per mezzo di un software *open source* di facile installazione. Le LOCKSS box interagiscono reciprocamente per controllare e confrontare, tramite il calcolo di *digest*, l'integrità degli oggetti digitali comuni avviando processi di riallineamento quando individuano delle differenze. In tal modo non occorrono i salvataggi perché se un appliance si accorge di avere dei contenuti danneggiati o incompleti li recupera interagendo con gli altri, secondo una logica molto simile alla condivisione *peer-to-peer*. Questa soluzione è stata concepita per mettere al centro dell'attività di conservazione le biblioteche, garantendo la sostenibilità economica ed organizzativa senza obbligarle ad impegni economicamente onerosi in termini di infrastruttura tecnica e di risorse umane necessarie alla gestione di tale processo.

Anche Portico e PMC hanno previsto evoluzioni verso il modello distribuito. Portico dichiara che nel 2009 supporterà la gestione di copie on-line su altri sistemi, attuando strategie di ridondanza e disaster recovery comunemente previste da molti data center. La NLM ha invece promosso PMC International (PMCI) per creare una rete di archivi digitali in grado di condividere i propri contenuti. Per la gestione degli archivi locali NCBI ha sviluppato il pacchetto software Portable PMC (pPMC) che opera su piattaforma Microsoft Windows e richiede Microsoft SQL Server e che comprende tutti i moduli funzionali necessari per l'importazione dei contenuti da PMC, il caricamento nel database, la ricerca, la riproduzione e la visualizzazione dei contenuti e per la gestione e l'elaborazione di statistiche.

Al momento non esistono esperienze e dati per consentire una comparazione tra il modello centralizzato e quello distribuito. Per esempio non è certo quale sia il tempo necessario, in funzione della banda di rete disponibile, della quantità e della dimensione degli oggetti digitali, per

ripristinare interamente i contenuti di una LOCKSS box in caso di crash e perdita totale dei dati archiviati, visto che non sono previste politiche di backup locale. D'altra parte occorre osservare che questi modelli sono fortemente correlati alle strutture organizzative ed agli investimenti necessari. Le soluzioni centralizzate comportano costi elevati in termini di risorse hardware ridondanti e affidabili e di risorse umane da impiegare per garantire adeguati livelli di sicurezza attraverso l'attuazione di politiche di salvataggio e delocalizzazione delle copie effettuate. Infatti Portico si propone come servizio terza parte a cui biblioteche ed editori affidano il compito della preservazione a lungo termine delle riviste elettroniche. Analogamente e-Depot è stato concepito nel contesto di una biblioteca nazionale che gestisce i servizi di deposito legale per il proprio paese. Per garantire il giusto equilibrio tra i requisiti di sostenibilità, sicurezza ed efficienza e per mantenere alle istituzioni ed alle biblioteche il controllo diretto dell'attività di conservazione delle riviste, al momento la soluzione più appropriata consiste nel creare una "rete a maglia di archivi affidabili" in grado di comunicare tra loro in modo sicuro e di condividere i contenuti dei propri *repository*. In tal modo, ogni singolo archivio può conservare una parte dei contenuti delle riviste elettroniche ottimizzando le attività di archiviazione e di gestione, mentre l'intero sistema (rete di archivi) è in grado di garantire alla comunità la copertura completa dei titoli da conservare e di assicurare un'adeguata ridondanza delle copie memorizzate.

## 5. Strategie per la conservazione

Uno dei principali fattori di rischio che può compromettere nel tempo la capacità di usare un oggetto digitale è la mancanza di informazioni su come interpretarlo e l'indisponibilità degli strumenti adatti a visualizzarlo o comunque a presentarlo in una forma comprensibile per la comunità dei potenziali utilizzatori. Il progetto InterPARES 1 [InterPARES] ha definito le strategie di digital preservation come un "*Coherent set of objectives and methods for maintaining digital components and related information over time, and for reproducing the related authentic records and/or archival aggregations*". I progetti di conservazione per affrontare i problemi legati all'obsolescenza e al cambiamento dei formati hanno adottato due strategie: la "migrazione" e la "emulazione".

La migrazione comprende il trasferimento dei contenuti digitali da un supporto all'altro, in base all'evoluzione dell'hardware e la conversione dei formati da quelli ormai obsoleti e non più supportati ai formati correnti. Queste trasformazioni possono avvenire o quando si rendono necessarie (e.g. PubMed Central) oppure al momento dell'accesso alla risorsa, tramite opportuni convertitori attivati in funzione delle caratteristiche operative dell'utente che sta accedendo al contenuto (e.g. LOCKSS). In alcuni casi una prima migrazione viene applicata nella fase precedente al caricamento nell'archivio concordando con i produttori i formati accettati per la fornitura dei contenuti (e.g. e-Depot, PubMed Central) oppure scegliendo di accettare una pluralità di formati che poi vengono trasformati al momento del caricamento (Portico, PubMed Central). In ciascuno di questi approcci i contenuti da archiviare vengono normalizzati in modo da gestire un numero limitato di formati standard, di solito PDF e XML, per ridurre la complessità del mantenimento nel tempo di troppi formati diversi.

L'emulazione è invece una tecnica mediante la quale la sequenza binaria (*bitstream*) originale viene memorizzata e mantenuta inalterata garantendone l'utilizzo (e.g. la visualizzazione) attraverso la simulazione dell'ambiente operativo originario per mezzo di opportuni software (si tratta di una situazione analoga a far girare un emulatore Linux su un PC Windows o un emulatore Windows su un PC Macintosh). La Koninklijke Bibliotheek (KB) nell'ambito dello sviluppo di e-Depot ha realizzato nel 2004 con l'IBM il prototipo dell' Universal Virtual Computer (UVC). Si tratta di una macchina virtuale che può essere implementata in modo indipendente dalla tecnologia sottostante. Il prototipo è stato realizzato in linguaggio Java. Su questa macchina vengono eseguiti programmi, scritti mediante opportuni strumenti di sviluppo, in grado di interpretare un qualunque formato

digitale di cui si conoscano le specifiche (il prototipo prevede i formati JPEG e GIF87a) che viene trasformato nel Logical Data View (LDV) che è un formato interno che può essere visualizzato tramite un opportuno programma. La KB sta proseguendo la sua attività di ricerca e sperimentazione nell'area dell'emulazione con il progetto Dioscuri.

Un confronto tra i costi delle strategie di migrazione e di emulazione è stato effettuato nell'ambito del progetto e-Depot [Oltmans&Koll] ed ha posto in evidenza che la migrazione è un processo applicato ripetitivamente a tutti gli oggetti della collezione, mentre l'emulazione richiede uno sviluppo iniziale valido per l'intera collezione. Considerato il costo una tantum per lo sviluppo dell'emulatore e che nel lungo periodo si devono riprodurre solo piccole porzioni delle collezioni, lo studio valuta che l'emulazione è più adatta per le collezioni molto ampie: per un archivio di cinque milioni di oggetti digitali e su un periodo di cinquanta anni viene stimato che i costi per la migrazione siano più del doppio di quelli per l'emulazione.

Bisogna sottolineare che ogni strategia di conservazione, dalla copia su un altro supporto fisico al cambiamento di formato, dalla normalizzazione in un formato standard alla visualizzazione della sequenza binaria originale attraverso un emulatore, comporta la creazione di copie e la manipolazione dell'oggetto digitale. È quindi sempre necessario valutare preventivamente la legittimità giuridica di queste operazioni sulla base e nel rispetto delle norme di legge sul diritto d'autore e dei contratti che regolano i rapporti con gli editori delle risorse [Coyle].

## 6. Indicazioni “pratiche” del rapporto CLIR

Il rapporto del CLIR [METES] formula una serie di raccomandazioni di carattere politico, organizzativo e operativo rivolte alle organizzazioni e biblioteche accademiche, agli editori e ai gestori di programmi di archiviazione delle riviste elettroniche e fornisce alcuni importanti strumenti di lavoro molto utili per tutti gli attori interessati alla tematica della conservazione degli *e-journal*, in primo luogo le istituzioni e le biblioteche: gli indicatori di valutazione e l'insieme minimo dei servizi.

### 6.1 Indicatori per la valutazione

L'esame condotto sui dodici programmi di conservazione delle riviste elettroniche si è stato basato sulla rilevazione e la comparazione di sette indicatori:

1. avere sia un'esplicita missione sia il necessario mandato per realizzare l'archiviazione nel lungo termine degli *e-journal*
2. negoziare tutti i diritti e le responsabilità necessari per adempiere ai propri obblighi per periodi lunghi;
3. descrivere esplicitamente quali pubblicazioni vengono archiviate e per conto di chi;
4. offrire un insieme di servizi di archiviazione ben definiti;
5. consentire l'accesso alle biblioteche in determinate condizioni;
6. dimostrare di essere vitali dal punto di vista organizzativo e operativo;
7. operare come parte di una rete.

Gli indicatori relativi alla missione, al mandato, all'accesso e alla vitalità organizzativa sono stati riscontrati per tutti i programmi censiti. La verifica sulle modalità di gestione dei diritti e delle responsabilità ha fornito risultati parziali. È stato accertato che in sette programmi non viene fornita adeguata informazione sulle riviste archiviate, mentre le informazioni raccolte sui servizi offerti sono state ritenute insufficienti per tutti i programmi. La cooperazione in rete è prevista per un solo progetto e parzialmente per altri sei. Solo cinque archivi (KB e-Depot, kopal/DDB, LOCKSS Alliance, LANL-RL, Portico) hanno espresso con certezza l'intenzione di diventare *trusted repository* certificato.

Questi indicatori forniscono alle biblioteche e alle istituzioni una metrica per la valutazione ed il confronto di servizi offerti da operatori esterni e le linee guida da seguire per la gestione di programmi di conservazione delle riviste elettroniche nel caso fossero intenzionate ad avviare autonomamente o in modo cooperativo iniziative di questo genere.

## 6.2 L'insieme minimo dei servizi

Il rapporto riprende ed amplia l'insieme dei servizi già individuati dalla "Urgent Action..." ponendo una particolare enfasi sull'importanza di implementare adeguate politiche di sicurezza:

a) *Ricevere i file che costituiscono la pubblicazione di una rivista in un formato standard, sia da una biblioteca partecipante che direttamente dall'editore.* In quest'ambito rientra la selezione dei componenti della rivista, la scelta di quali formati accettare (quelli originali forniti dai produttori oppure un formato standard, di norma il PDF) e la decisione in merito al mantenimento delle caratteristiche grafiche.

b) *Memorizzare i file in formati non proprietari che possano essere facilmente trasferiti e usati nel caso in cui la biblioteca dovesse decidere di cambiare i propri archivi.* Il rapporto sottolinea l'importanza di adottare metadati e formati standard.

c) *Usare metodi standard per verificare l'integrità dei file (in ingresso ed in uscita) e effettuare continuamente dei controlli d'integrità sui file memorizzati internamente.* Le strategie di conservazione, da applicare fin dalla fase iniziale di acquisizione dei contenuti, devono comportare la periodica verifica della correttezza dei dati archiviati per assicurarne l'integrità.

d) *Limitare il trattamento dei file ricevuti per contenere i costi, ma effettuare le elaborazioni necessarie in modo che in caso di perdita gli archivi possano adeguatamente localizzare e riprodurre i file per le biblioteche aderenti al servizio.* Quest'area comprende i processi per il caricamento (*ingest*) dei contenuti nell'archivio, inclusi i controlli di integrità e completezza dei file forniti e la creazione dei metadati.

e) *Attivare adeguate protezioni contro la perdita dovuta a danni fisici per mezzo della memorizzazione ridondante e le comuni e ben documentate misure di sicurezza.* Si tratta di attivare le adeguate misure di sicurezza per prevenire la perdita o l'alterazione dei dati (accidentale o dolosa) agendo su tutte le componenti coinvolte nel processo di archiviazione: tecniche, logistiche, organizzative e finanziarie. Lo studio ha anche evidenziato che l'utilizzo attivo e continuo dell'archivio (in contrasto con l'approccio dark archive) è un metodo complementare a quello dei test per verificare l'integrità dei contenuti.

f) *Offrire strumenti aperti e trasparenti per verificare e certificare le suddette pratiche.* È fondamentale che le politiche, i processi e le attività in cui si articola la *digital preservation* siano accompagnati da adeguata documentazione, costantemente aggiornata e consultabile da parte delle comunità di riferimento, e sottoposti a continua valutazione.

Tutti questi servizi sono facilmente riconducibili al modello funzionale astratto definito da OAIS, ma il rapporto mette in evidenza che solo cinque delle iniziative censite (Canada Institute for Scientific and Technical Information CSI, KB e-Depot, kopal/DDB, LOCKSS Alliance, Portico) hanno dichiarato la propria aderenza allo standard OAIS. In particolare dalla documentazione disponibile sul sito web la conformità di LOCKSS risulta abbastanza approssimativa. Le raccomandazioni formulate non comprendono la conformità al modello OAIS, prescritta invece dal

rapporto OCLC-RLG sui *trusted repository*, probabilmente per la forte sottolineatura di “limitare il trattamento dei file ricevuti per contenere i costi”. Tuttavia si ritiene che l’adozione dei modelli OAIS e la loro implementazione attraverso standard consolidati e aperti (metadati, formati, identificatori, etc.) siano due principi difficilmente derogabili per realizzare archivi affidabili che garantiscano la conservazione delle riviste elettroniche. A tal fine sono di notevole utilità gli studi svolti e la documentazione prodotta dal progetto europeo CASPAR.

## 8. Conclusioni: alla ricerca di punti fermi

Le biblioteche dedicano una parte sempre più significativa dei loro investimenti alla sottoscrizione di contratti per l’accesso alle riviste in formato elettronico, spesso attraverso la partecipazione a consorzi appositamente costituiti. Il passaggio irreversibile dagli abbonamenti che consentono di acquisire le copie cartacee delle riviste alla stipula di licenze che abilitano l’accesso remoto al formato elettronico pone in primo piano una serie di problemi legati alla disponibilità futura di queste risorse. In particolare occorre individuare quali azioni intraprendere per assicurare che tali contenuti siano utilizzabili anche dopo la scadenza dei contratti d’uso e dopo che l’editore abbia dismesso una rivista, cessato la propria attività o quando si siano verificati problemi tecnici che impediscono per un periodo prolungato l’accesso alle sue pubblicazioni.

Dopo il pressante appello lanciato alla comunità accademica, scientifica e bibliotecaria dalla “Urgent Action Needed to Preserve Scholarly Electronic Journals” il rapporto CLIR ha messo in risalto che gli attuali tipi di contratti non forniscono da parte degli editori alle biblioteche e ai loro utenti adeguate garanzie per l’accesso prolungato nel tempo ai contenuti delle riviste elettroniche. Quindi occorre che le biblioteche, la comunità accademica e scientifica e le istituzioni culturali promuovano iniziative volte a garantire la conservazione nel tempo delle riviste digitali attraverso lo sviluppo di opportuni servizi ed un confronto serrato con gli editori e gli aggregatori sia sulle politiche commerciali che sulle garanzie offerte per l’uso prolungato nel tempo.

Inoltre le recenti evoluzioni relative alla composizione ed all’articolazione dei contenuti delle riviste digitali richiedono lo sviluppo di progetti di *digital preservation* adeguatamente concepiti per garantire il mantenimento delle citazioni bibliografiche, dei riferimenti ai data set e degli oggetti generati dinamicamente o esterni.

Le iniziative analizzate mostrano che ancora non esiste un modello condiviso e consolidato sia sul piano tecnico che su quello organizzativo. Tuttavia l’analisi delle pratiche adottate permette di affermare che un programma avente per oggetto la conservazione delle riviste elettroniche deve comunque tener presenti almeno tre punti di riferimento: a) la realizzazione di *trusted repository*; b) l’adozione dei modelli informativo e funzionale di OAIS e c) l’applicazione delle metodologie di valutazione e di verifica. Si tratta di tre capisaldi necessari per garantire alla comunità di riferimento l’affidabilità e la continuità del servizio offerto.

Sul piano organizzativo, considerati gli elevati costi implicati dalla implementazione e gestione dei servizi di conservazione del digitale, è ampiamente condivisibile l’affermazione di [Oltmans&Van Wijngaarden] “*Not every library should try to establish its own permanent archiving system. In the case of international scholarly journals a handful of permanent archives, widely spread around the globe, will suffice. The economies of scale that can be achieved provide a key incentive for developing this safe place model*”. Nessuna istituzione o biblioteca può affrontare con le sole proprie forze questa sfida. La cooperazione e la condivisione sono l’unico approccio che possa permettere di sostenere la complessità ed i costi dei processi di conservazione. Pertanto si auspica, attraverso la sensibilizzazione delle istituzioni (Regioni, Università, MUR, MIBAC) e degli organismi (CRUI, Comitato Nazionale SBN) più direttamente coinvolti, la creazione di una rete nazionale di archivi affidabili distribuiti. Tale infrastruttura dovrebbe essere sviluppata e gestita secondo un modello fortemente cooperativo, avvalendosi del supporto di gruppi di studio (AIB,INFER) e dei consorzi nazionali (CASPUR, CILEA, CIPE) già esistenti e della partecipazione

alle iniziative di livello internazionale (e.g. CLOCKSS, PORTICO), con l'obiettivo di garantire un'adeguata copertura e la preservazione dei contenuti delle riviste elettroniche per le future generazioni.

## 9. Bibliografia

[Chandra] Chandra Prabha. "*Shifting from Print to Electronic Journals in ARL University Libraries*" *Serials Review* 33, no. 1, March 2007

[Coyle] Karen Coyle. "*Rights in the PREMIS Data Model.A*" report for the Library of Congress, Library of Congress, December 2006.

URL: <http://www.loc.gov/standards/premis/Rights-in-the-PREMIS-Data-Model.pdf>

[CRL] *Center for Research Libraries Certification of Digital Archives* home page.

URL: <http://www.crl.edu/content.asp?l1=13&l2=58&l3=142>

[Crow] Raym Crow. "*Publishing Cooperatives: An Alternative for Society Publishers*". Scholarly Publishing and Academic Resources Coalition, February 2006.

URL: [http://www.arl.org/sparc/bm%7Edoc/Cooperatives\\_v1.pdf](http://www.arl.org/sparc/bm%7Edoc/Cooperatives_v1.pdf)

[DCC] *Digital Curation Centre* home page. URL: <http://www.dcc.ac.uk/>

[DRAMBORA] *Digital Repository Audit Method Based on Risk Assessment (DRAMBORA)* toolkit.

URL: <http://www.repositoryaudit.eu/>.

[InterPARES] *The International Research on Permanent Authentic Records in Electronic Systems (InterPARES)*.

URL: <http://www.interpares.org/>

[JISC] *JISC Repositories and preservation programme* home page.

URL: [http://www.jisc.ac.uk/whatwedo/programmes/programme\\_rep\\_pres.aspx](http://www.jisc.ac.uk/whatwedo/programmes/programme_rep_pres.aspx)

[Johnson&Luther] Richard K. Johnson and Judy Luther. "*The E-only Tipping Point for Journals. What's Ahead in the Print-to-Electronic Transition Zone*", Association of Research Libraries (ARL), Washington, 2007.

URL: [http://www.arl.org/bm~doc/Electronic\\_Transition.pdf](http://www.arl.org/bm~doc/Electronic_Transition.pdf)

[Lynch] Clifford Lynch. "*The Shape of the Scientific Article in The Developing Cyberinfrastructure*", CTWatch Quarterly, August 2007.

URL: <http://www.ctwatch.org/quarterly/articles/2007/08/the-shape-of-the-scientific-article-in-the-developing-cyberinfrastructure/>

[METES] Anne R. Kenney, Richard Entlich, Peter B. Hirtle, Nancy Y. McGovern, and Ellie L. Buckley. "*E-Journal Archiving Metes and Bounds: A Survey of the Landscape*", Council on Library and Information Resources, Washington, D.C. September 2006.

URL: <http://www.clir.org/PUBS/abstract/pub138abst.html>

[MINIMUM-CRITERIA] *Minimum Criteria for an Archival Repository of Digital Scholarly Journals*, Digital Library Federation(DLF), version 1.2, May 2000.

URL: <http://www.diglib.org/preserve/criteria.htm>

[Mogge] Dru Mogge, editor. *Directory of Electronic Journals, Newsletters, and Academic Discussion Lists*, 7th edition. Washington, D.C.: Association of Research Libraries. 1997.

URL:

<http://www.emeraldinsight.com/Insight/html/Output/Published/EmeraldFullTextArticle/Pdf/2380170103.pdf>

[NESTOR] *nestor - studies 8 - Catalogue of Criteria for Trusted Digital Repositories - Version 1*, nestor Working Group Trusted Repositories – Certification, Frankfurt am Main : nestor c/o Deutsche Nationalbibliothek, 2007. URL: <http://nbn-resolving.de/urn:nbn:de:0008-2006060703>

[NEWJOUR] *the archive for New Jour, the Internet list for new journals and newsletters available on the Internet*, University of California at San Diego Libraries (UCSD).

URL: <http://gort.ucsd.edu/newjour/>

[OAIS] CCSDS 650.0-B-1: *Reference Model for an Open Archival Information System (OAIS)*. Blue Book (Standard ISO 14721:2003). Issue 1. January 2002.

URL: <http://public.ccsds.org/publications/archive/650x0b1.pdf>

[Oltmans&Kol] Erik Oltmans & Nanda Kol. “*A Comparison Between Migration and Emulation in Terms of Costs*”, RLG DigiNews April 15, 2005.

URL: <http://digitalarchive.oclc.org/da/ViewObject.jsp?objid=0000068902&reqid=18109#article0>

[Oltmans& van Wijngaarden] Erik Oltmans and Hilde van Wijngaarden. “*The KB e-Depot digital archiving policy*”, Koninklijke Bibliotheek, The Hague, The Netherlands Library Hi Tech, Vol. 24 No. 4, 2006. pp. 604-613.

URL:

<http://www.emeraldinsight.com/Insight/ViewContentServlet?Filename=Published/EmeraldFullTextArticle/Articles/2380240412.html>

[Pearce-Moses] Richard Pearce-Moses. “*A Glossary of Archival and Records Terminology*”, The Society of American Archivists, Chicago, 2005. URL: <http://www.archivists.org/glossary/index.asp>

[PREMIS] *PREMIS (PREservation Metadata: Implementation Strategies) Working Group* home page. URL: <http://www.oclc.org/research/projects/pmwg/default.htm>

[RLG-OCLC] “*Trusted Digital Repositories: Attributes and Responsibilities*”, RLG Mountain View, CA, May 2002

URL: <http://www.oclc.org/programs/ourwork/past/trustedrep/default.htm>

[TRAC] “*Trustworthy Repositories Audit & Certification (TRAC): Criteria and Checklist*”, revised and expanded version of The Audit Checklist for the Certification of Trusted Digital Repositories, originally developed by RLG-NARA Digital Repository Certification Task Force, OCLC and CLR, Vers. 1, February 2007.

URL: <http://www.crl.edu/content.asp?l1=13&l2=58&l3=162&l4=91>

[URGENT-ACTION] “*Urgent Action Needed to Preserve Scholarly Electronic Journals*”, Edited by Donald J. Waters, The Andrew W. Mellon Foundation, October 2005.

URL: <http://www.diglib.org/pubs/waters051015.htm>

## Progetti Europei

CASPAR. *Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval.*

URL: <http://www.casparpreserves.eu/>

DIOSCURI. *Dioscuri - the modular emulator.* URL: <http://dioscuri.sourceforge.net/>

DPE. *DigitalPreservationEurope (DPE).* URL: <http://www.digitalpreservationeurope.eu/>

JISC Repositories and preservation programme.

URL: [http://www.jisc.ac.uk/whatwedo/programmes/programme\\_rep\\_pres.aspx](http://www.jisc.ac.uk/whatwedo/programmes/programme_rep_pres.aspx)

NESTOR *the german Network of expertise in Digital long-term preservation.*

URL: [http://www2.hu-berlin.de/nesstor/nesstor\\_en/index.php](http://www2.hu-berlin.de/nesstor/nesstor_en/index.php)

PLANETS. *Preservation and Long-term Access through Networked Services.*

URL: <http://www.planets-project.eu/>

Link controllati il 05 Gennaio 2008