

Preservazione delle pubblicazioni scientifiche digitali e depositi istituzionali: problematiche e tendenze in Italia

di Anna Maria Tamaro, Università di Parma

Anche se il loro sviluppo è recente, i depositi istituzionali sono da considerare un importante fenomeno per la comunicazione scientifica in ambito digitale. La preservazione nei depositi istituzionali si intende non limitata alla preservazione dei supporti fisici di memorizzazione ma è inclusiva della pianificazione, dello stanziamento delle risorse e nell'applicazione dei metodi e delle tecnologie per la conservazione preventiva, necessari ad assicurare che le pubblicazioni scientifiche digitali di valore duraturo memorizzate nei depositi rimangano accessibili e utilizzabili. I depositi istituzionali dovrebbero considerare la preservazione come loro funzione prioritaria per consentire l'accesso permanente alle pubblicazioni scientifiche digitali, tuttavia la preservazione non sempre è stata considerata con la necessaria attenzione e in genere si è ancora ben lontani dall'aver chiarezza su come realizzarla.

La presentazione intende analizzare come è attuata in Italia l'organizzazione della preservazione nei depositi istituzionali. Molti sono i problemi, tecnici, organizzativi, giuridici e economici che devono essere affrontati. L'analisi terrà conto dell'approccio organizzativo (digital curation), dei formati garantiti, degli identificatori selezionati e, soprattutto, dei costi, insieme alla possibilità di usare la cooperazione con altre istituzioni per la sostenibilità nel tempo del servizio.

1. Introduzione

La presentazione si concentra sulla preservazione delle pubblicazioni scientifiche digitali. Con l'avvento di Internet e del Web, molte delle pubblicazioni scientifiche sono accessibili attraverso la rete, come prodotto dell'attività di conversione digitale e come prodotti originari dell'editoria digitale. Questo ha creato una serie di problematiche, giuridiche, tecnologiche, economiche ed organizzative, che vogliamo brevemente descrivere.

Dobbiamo prima di tutto chiederci: cosa è una pubblicazione scientifica?

1.1 Definizione di pubblicazione scientifica digitale

La definizione legale di «pubblicazione scientifica» intende non una qualsiasi riproduzione a stampa, ma l'opera pubblicata da un editore. Lo ha stabilito il Consiglio di Stato, con la sentenza n. 2364 del 22 aprile 2004, precisando che la diffusione fra il pubblico della pubblicazione è un requisito necessario di cui l'opera «edita» deve essere in possesso (ma soltanto se stampata in Italia). La definizione legale di pubblicazione scientifica prevede tre elementi necessari: 1) deve essere un'opera dell'ingegno 2) deve essere riprodotta in forma tangibile e 3) deve essere divulgata al pubblico da un editore, che è l'operatore il quale cura non soltanto la riproduzione a stampa di un'opera, ma la sua diffusione fra il pubblico. Dal punto di vista della preservazione, la pubblicazione scientifica in formato cartaceo è facilitata dalla molteplicità di copie, ed è certificata dalla sua presenza in una biblioteca nazionale, mediante il deposito legale.

Questa definizione legale non soddisfa i requisiti della pubblicazione scientifica digitale: anche questa è un'opera dell'ingegno, ma di cui è a volte difficile individuare l'autore, è caratterizzata da flessibilità invece che della stabilità di una forma fisica, ed infine è diffusa attraverso la rete e può non avere un editore.

La flessibilità è il primo elemento che caratterizza la pubblicazione scientifica digitale. Una pubblicazione digitale è instabile, dinamica, ipertestuale e senza confini definiti, se non l'architettura delle banche dati che la contengono, distribuite in rete. Questo ha varie conseguenze nell'autenticazione e localizzazione della pubblicazione digitale. Le sequenze di bit possono soffrire di una facile falsificazione o anche di una perdita non voluta di informazioni. Senza adeguate misure preventive un'alterazione del contenuto potrebbe non lasciare tracce. Differenti versioni del documento ipertestuale vengono create continuamente, con una conseguenza: quello che è difficile da definire è anche difficile da descrivere e quindi da organizzare. I metadati, elementi descrittivi possono essere incapsulati nei documenti stessi ma tutti i documenti devono essere descritti in modo simile, prescindendo dal supporto, proprio per garantire l'integrazione e l'interoperabilità funzionale in rete di documenti dinamici. Per la certificazione delle pubblicazioni digitali occorre quindi una soluzione che garantisca l'integrità nel tempo di una pubblicazione digitale (assenza di qualsiasi alterazione, contenuto non modificato, preservazione nel tempo di una versione citabile), insieme alla descrizione e identificazione univoca composta in un certo momento da un determinato soggetto.

Un altro elemento caratterizzante le pubblicazioni scientifiche digitali è l'auto-pubblicazione, cioè la possibilità di diffondere una pubblicazione in rete senza l'opera di un editore. La tecnologia consente a tutti di diventare autore e di essere editori di se stesso, nel senso di

poter diffondere una pubblicazione scientifica digitale attraverso Internet. Il fenomeno dell'auto-pubblicazione ha riguardato in un primo tempo la pubblicazione in linea parallela a quella diffusa nella tradizionale editoria cartacea, come articoli di riviste (e-print), libri elettronici, e documenti multimediali. Ci sono tuttavia anche nuove tipologie di pubblicazioni scientifiche, che non hanno analogie con documenti su carta. Ad esempio, possiamo elencare i siti web, come "Reti medievali", in cui si può avere accesso a grosse quantità di dati di ricerca, ed a diverse tipologie di informazioni di qualità. Potremo inserire in questo elenco di nuove pubblicazioni anche le tipologie di comunicazione informale, come le liste di discussione, i blog, le chat, i RSS feeds, le collezioni di collegamenti (bookmarks): tutte queste pubblicazioni digitali, anche se informali, contengono una quantità di fatti, discussioni e rapporti che possono far parte della comunicazione scientifica e possono essere memorizzati, poiché ogni comunicazione digitale lascia una traccia. Naturalmente, questa possibilità di esprimere la propria creatività con l'auto-pubblicazione di pubblicazioni scientifiche in forma tradizionale o nuova, può lasciare a ridere in quanto a qualità ed accuratezza. Non essendoci il valore aggiunto della peer review, effettuata dagli editori, in molti casi la pubblicazione nel Web è considerata un caso di "vanity press", di scarso valore scientifico. Quando gli autori hanno cominciato a mettere in rete i loro lavori, il contrasto con gli editori è stato inevitabile. In una prima fase, si è diffusa la "regola di Ingelfinger": Ingelfinger era il direttore della rivista *New England Journal of Medicine*, e la regola con il suo nome dice che il periodico può pubblicare solo lavori originali e non pubblicati prima.

La situazione di attrito con gli editori ora è cambiata con il movimento Open Access e lo sviluppo degli archivi e dei depositi istituzionali. Il movimento Open Access è nato intorno alla metà degli anni '90, partendo dalle iniziative di auto-pubblicazione di alcuni autori accademici, a cui poi l'istituzione universitaria o di ricerca ha deciso di dare una infrastruttura di supporto a livello istituzionale. I depositi istituzionali si pongono in questa linea di tendenza e costituiscono uno dei componenti più importanti dell'infrastruttura per la nuova comunicazione scientifica. Ci sono molte definizioni di Open Access, alcune sono larghe, altre più specifiche, ma l'elemento comune è che l'utente finale ha accesso libero e gratuito ai contenuti digitali. Il modello Open Access che è prevalente vede due responsabilità che sono distribuite: quella dei fornitori di contenuti e metadati e quella dei fornitori di servizi. Ma il problema è: chi è responsabile della preservazione di lungo periodo?

Uno dei paradossi del digitale è che, pur se le tecnologie favoriscono la produzione diffusa di contenuti digitali e facilitano l'accesso, mancano ancora istituzioni e procedure che siano responsabili di certificare la pubblicazione digitale, organizzarla e preservarla nel tempo. Il problema delle pubblicazioni scientifiche in formato digitale non va quindi limitato al

cambiamento di supporto dal cartaceo al digitale; coinvolge invece il concetto tradizionale di pubblicazione, che è legato alla stabilità del supporto fisico ed ad un processo editoriale in un flusso che nei secoli è stato ben definito e che garantisce alle pubblicazioni la loro certificazione e la loro identificazione e localizzazione permanente.

Immaginiamo che nel 2060, qualcuno vorrà recuperare una pubblicazione scientifica in formato digitale che trova citata in un testo. Come abbiamo anticipato, poiché le pubblicazioni scientifiche digitali non sono certificate, una serie di ostacoli e problemi possono essere anticipati per il singolo ricercatore:

- non potrà affidarsi a strumenti bibliografici di reperimento, come i cataloghi delle biblioteche, le bibliografie tematiche, altri strumenti: le biblioteche e le bibliografie hanno rinunciato a descrivere tutto quello che viene prodotto nelle università e nelle istituzioni di ricerca, adottando filtri più o meno stretti per la selezione e la catalogazione delle pubblicazioni scientifiche;
- non potrà affidarsi ai motori di ricerca, presupponendo che anche in futuro, la ricerca sia basata su questi strumenti: le pubblicazioni scientifiche sono spesso gestite da sistemi proprietari nel cosiddetto “deep web”, che non viene indicizzato dai motori di ricerca;
- non potrà trovare la pubblicazione depositata presso la biblioteca nazionale, o la biblioteca dell’università di riferimento dell’autore: gli strumenti tradizionali di deposito delle pubblicazioni non sono estesi alle pubblicazioni digitali ed il deposito legale, se c’è, è volontario;
- non potrà ritrovare la pubblicazione alla localizzazione indicata (URL), che potrà essere stata cambiata, senza che la pubblicazione citata abbia avuto un identificatore univoco ed una localizzazione persistente, ricevendo il comune “Error 404” in risposta, cosa abbastanza diffusa anche dopo solo poco tempo dalla messa in rete della pubblicazione;
- non potrà quindi citare una pubblicazione scientifica, che non sarà più identificabile, ad esempio una volta che il sito Web o il blog dell’autore è stato abbandonato, o non più raggiungibile.

Gli sforzi per costruire un’infrastruttura che sia allo stesso tempo capace di certificare e preservare le pubblicazioni digitali stanno però guadagnando terreno. Possono essere ad esempio evidenziati in Italia gli sforzi congiunti della Biblioteca Nazionale di Firenze, della Biblioteca del CNR di Roma, delle biblioteche universitarie: tutte queste istituzioni stanno cercando di realizzare il deposito legale delle pubblicazioni digitali ed il connesso problema della descrizione bibliografica e della certificazione della versione depositata. Posso citare ad

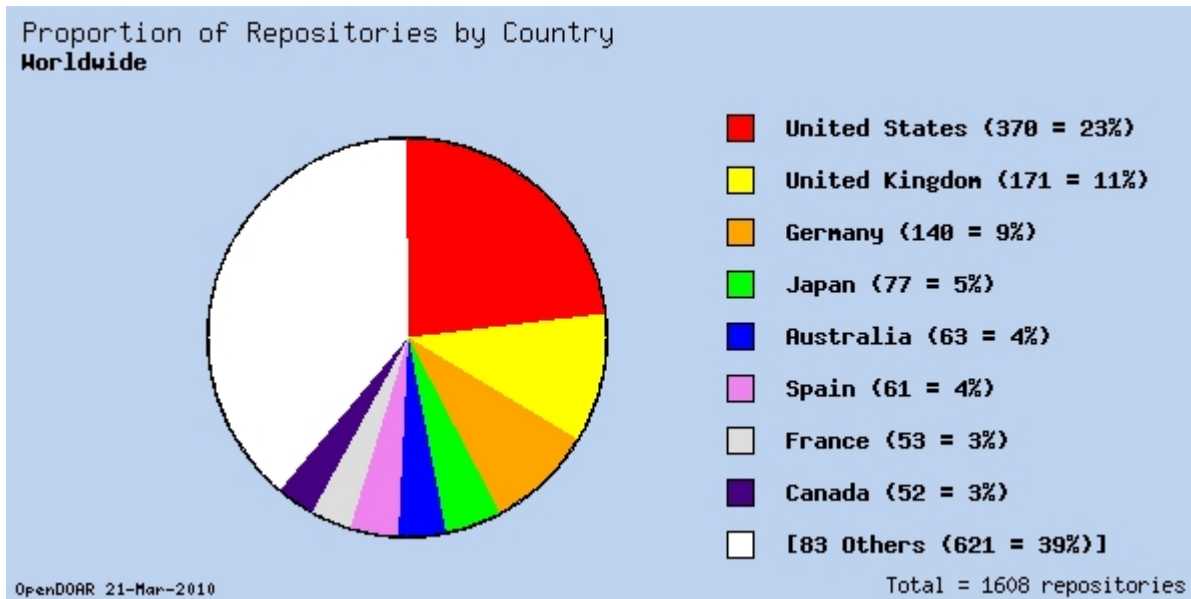
esempio il progetto che è stato finanziato dalla Fondazione Rinascimento Digitale dal titolo “Magazzini Digitali”, ed il progetto dell’identificatore univoco NBN (National Bibliographic Number) che è stato adottato come identificatore univoco delle pubblicazioni depositate. Questi progetti pilota, nel ritardo della legge nazionale sul deposito legale delle pubblicazioni digitali, hanno ottenuti significativi risultati pratici, sono in avanzata fase di sperimentazione e possiamo dire che stanno anticipando la legislazione stessa, che ora potrà basarsi sulla prassi che è stata stabilita a livello volontario dalle singole istituzioni. Tra le buone pratiche che cercano di creare un’infrastruttura per le pubblicazioni scientifiche digitali, i depositi istituzionali sono una delle esperienze più interessanti e su questi vogliamo concentrarci.

2. Depositi istituzionali

Una delle più comuni definizioni di deposito istituzionale è quella secondo cui (Branin 2004) questi sono depositi dei contenuti digitali prodotti dall’istituzione scientifica, con la funzione di memorizzare, gestire, preservare e renderli accessibili nel Web. I depositi che sono avviati sempre più spesso dalle università, costituiscono una componente essenziale dell’infrastruttura di ricerca che è necessaria per le pubblicazioni scientifiche digitali (Tammaro 2007; Tammaro 2008). A differenza degli archivi istituzionali, i depositi sono meno interessati all’auto-pubblicazione e concentrati nel garantire la funzionalità di accesso permanente ai risultati della didattica e della ricerca istituzionale. Gli scopi (Lynch 2003) per cui le università avviano un deposito istituzionale non sono limitati al recupero dell’informazione, ma si estendono a comprendere la trasparenza sui risultati di ricerca e la connessa valutazione di qualità dell’istituzione, che ricerca visibilità e riconoscimento.

In Italia abbiamo attualmente circa 51 depositi istituzionali, il 3% del totale (circa 1600) dei depositi nel mondo (Fig. 1), secondo le statistiche offerte da OpenDoar¹, che è il registro dei depositi istituzionali.

¹ OpenDoar è accessibile a: <<http://www.opendoar.org/>>. I depositi istituzionali sono indicizzati dai motori di ricerca che estraggono alcuni metadati dai dati bibliografici (sulla base del protocollo OAI-PMH) e possono inoltre essere interrogati da PLEAIDI, che è un portale che aggrega la ricerca sulle pubblicazioni scientifiche italiane in linea: <<http://www.openarchives.it/pleiadi/>>.



I depositi istituzionali con la maggiore quantità di contenuti sono presso l'Università di Milano (31.296 documenti comprendenti presentazioni a Conferenze, tesi, articoli e pre-print), l'Istituto Superiore di Sanita (25.675 documenti con presentazioni di Conferenze e pre-print), l'Istituto Universitario Europeo (6.780 documenti che includono articoli e pre-print), la FAO (260.441 documenti che includono articoli, presentazioni a Conferenze, pre-print).

Dalla comparazione con altri depositi istituzionali in OpenDoar, possiamo evidenziare che, mentre il numero di depositi istituzionali in Italia è equivalente a quello di altre nazioni europee, la quantità dei documenti raccolti è più limitata. Questo perché i depositi istituzionali non sono ancora considerati una strategia dell'istituzione e manca un mandato istituzionale chiaro (Bevilacqua 2008). La tipologia dei documenti raccolti comprende soprattutto e-proceedings di Conferenze, tesi e pre-print. Le pubblicazioni scientifiche, come libri ed articoli di periodici, soprattutto in aree come quella umanistica e sociale, restano ancora cartacee. Non vengono raccolti nei depositi istituzionali le nuove pubblicazioni come blog, liste, siti Web e neppure le banche dati che raccolgono i risultati di ricerca. Per le banche dati c'è una crescente domanda di memorizzazione nei depositi ma ci sono alcuni problemi da risolvere, come la quantità di solito numerosa dei dati da conservare, la necessità di conservare anche la loro struttura e non solo il contenuto, la necessità di ri-usare i dati, con funzionalità aggiunte che quindi dovrebbero essere offerte dal software del deposito.

Pur nei limiti di queste debolezze organizzative, i depositi istituzionali assicurano attualmente una funzione essenziale. C'è una differenza sostanziale tra le pubblicazioni scientifiche digitali che possono essere trovate nei siti Web e quelle che si trovano nei depositi istituzionali: nel 2060 solo le pubblicazioni scientifiche digitali nei depositi istituzionali potranno essere ricercate ed usate. Questa differenza è il risultato di due funzionalità che i

depositi assicurano: la “digital curation” e la preservazione. Senza queste due funzionalità non c’è nessuna garanzia di accesso permanente. La preservazione si può dire che è il processo passivo di conservazione nel tempo, mentre la “digital curation” è la procedura attiva: a differenza delle pubblicazioni su carta, le pubblicazioni digitali non possono essere preservate solo se lasciate in un deposito, ma richiedono invece una cura costante ed attiva di gestione.

2.1 Digital curation

Il concetto di cura digitale (chiamato diffusamente *digital curation*) individua un ruolo nuovo o rinnovato per molte istituzioni, soprattutto archivi, biblioteche e musei, ma anche i depositi istituzionali, di cui qui ci interessiamo. Questa attività è svolta da un professionista, meglio se esperto della materia, che seleziona, preserva, mantiene e gestisce un assetto digitale, per la ricerca presente e futura dei ricercatori. La *digital curation* comprende anche tutte le attività di conversione di formato, di migrazione su supporti diversi, di inserimento dei metadati, di tutte le procedure per assicurare l’autenticità ed integrità.

Questa attività di “digital curation” è diversa da quella svolta da Google, che al momento della domanda non fa altro che aggregare contenuti digitali; è diversa anche da quella svolta dagli editori, che creano le loro collezioni, anche se la “digital curation” si può configurare come un’attività editoriale di selezione e standardizzazione delle pubblicazioni digitali.

La “digital curation” implica da una parte definizioni certe per procedure, standard, responsabilità specifiche, incluso per le pubblicazioni scientifiche digitali la necessità di quelle certezze giuridiche che sono necessarie nella produzione e diffusione delle stesse. Dall’altra parte, quella dell’accesso permanente, la “digital curation” consente di rendere leggibili a distanza di tempo le pubblicazioni scientifiche e di gestirle correttamente, di controllarle e conservarle in modo opportuno per ritrovarne la funzionalità originaria.

Criteri di selezione

Un primo elemento di riflessione riguarda i criteri di selezione delle pubblicazioni digitali da conservare, quelle cioè che avranno un’utilità, un significato anche in futuro, ad esempio nel 2060. Alcune istituzioni accademiche hanno favorito un approccio archivistico, altre, considerando gli alti costi che la conservazione del digitale comporta, fanno una selezione legata alla funzionalità. Il sistema australiano AGLS, ad esempio, che memorizza documenti prodotti dal governo, prevede tra i metadati MAG un elemento (chiamato *validity date range*), in cui vengono registrate le date entro cui il documento è considerato valido. I documenti nella cui descrizione appare questo elemento saranno tenuti sotto controllo,

anche mediante automatismi di sistema, per la possibile eliminazione nel tempo, o comunque per essere destinati a trattamenti di conservazione di altro tipo

Metadati e formati

Per migliorare l'interoperabilità e la preservazione delle pubblicazioni scientifiche, c'è una rinnovata esigenza di norme standard. Queste sono attività legate a standard che riguardano aree come l'interazione tra produttore di contenuti e curatore dell'archivio, gli identificatori persistenti come componenti essenziali per la preservazione, metadati per la preservazione, formati per la preservazione (ad esempio, il formato PDF-A per i testi). Gli standard possono portare a risultati significativi per l'interoperabilità, in quanto codificano molte delle migliori pratiche che sono state identificate negli ultimi anni dai progetti pilota.

La preservazione è necessaria non solo per le collezioni digitali ma anche per i metadati ed i formati delle pubblicazioni digitali. I depositi istituzionali, come scelta razionale e per limitare i rischi di perdita di contenuti, usano limitare il numero dei formati che sono garantiti, come ad esempio il formato PDF o XML. Nei depositi istituzionali che utilizzano DSpace, ad esempio, oltre alla possibilità di utilizzare i metadati amministrativo-gestionali necessari, è disponibile uno strumento che ha lo scopo di rendere più agevole le operazioni di preservazione dei documenti, sotto il profilo del controllo dei formati: questo strumento è il registro dei formati (*bitstream registry*), in cui vengono registrati i formati per i quali l'istituzione si impegna a garantire un'operazione di conservazione perenne, con le tecniche che si riterranno più opportune anche riguardo alla tipologia dei documenti in questione (*refreshing*, emulazione, ecc.). Il creatore che introduce un documento per l'archiviazione, può prendere visione del registro, in modo da essere consapevole del "destino" futuro del proprio lavoro, ed eventualmente presentarlo in un formato diverso, scegliendo tra quelli per i quali è garantita la preservazione.

I depositi istituzionali, che sono nati nel contesto del movimento Open Access, adottano standard che sono sempre aperti e mai proprietari. L'apertura è essenziale per l'interoperabilità che deve essere garantita per la preservazione e che riguarda l'interoperabilità tra ambienti software, come anche l'interoperabilità dei formati di conservazione. I formati aperti, come quelli usati per i sistemi geografici (OpenGIS), o sistemi di ufficio (OpenOffice), ed anche strutture chimiche (Molfiles e SMILES), vanno verso l'indipendenza da hardware e software. Il potenziale per l'utilizzo di questi formati aperti per la conservazione è un'area di ulteriore ricerca.

Identificatori

Un altro strumento necessario per garantire l'accessibilità futura dei documenti è l'utilizzo degli identificatori univoci (Lynch 1998; Vitiello 2004): parliamo qui di identificatori che hanno una validità globale nel Web, gestiti da entità che si pongono a un livello superiore rispetto ai singoli depositi, e quindi di identificatori che sono cosa diversa da quelli attribuiti localmente dal sistema automatizzato. Nel modello Open Access, queste entità superiori sono agenzie pubbliche ed anche private ma in ogni caso no-profit, che attribuiscono gli identificatori, anche se le pubblicazioni scientifiche raccolte nei depositi istituzionali possono avere identificatori proprietari come il DOI.

Il DSpace, ad esempio, implementa l'*handle system*, un software che attribuisce automaticamente un numero identificativo nel momento in cui il documento è accettato e quindi va ad essere inserito nell'archivio. Questo numero è composto da una parte fissa, che è l'identificativo attribuito all'istituzione responsabile dell'archivio dal CNRI, agenzia che gestisce il sistema in modo gratuito, e dà un numero progressivo. Si tratta quindi di una sequenza numerica univoca che permane anche se la risorsa digitale viene trasferita in diversi depositi. Contemporaneamente il documento riceve un URI, che è formata dall'indirizzo IP del sistema di gestione centralizzato degli handle², più l'handle che identifica il documento. Infatti, in questo modo, l'handle resolver dell'agenzia no-profit CNRI sarà capace di localizzare il documento ovunque esso sia nella rete, indipendentemente da quella che è stata la sua storia (cambiamento di deposito, cambiamento dell'istituzione che lo gestisce, etc.). Nelle citazioni che saranno fatte del documento andrà utilizzata sempre questa URI per permettere anche in futuro di rendere rintracciabili i documenti citati, così come lo è ora per i documenti a stampa.

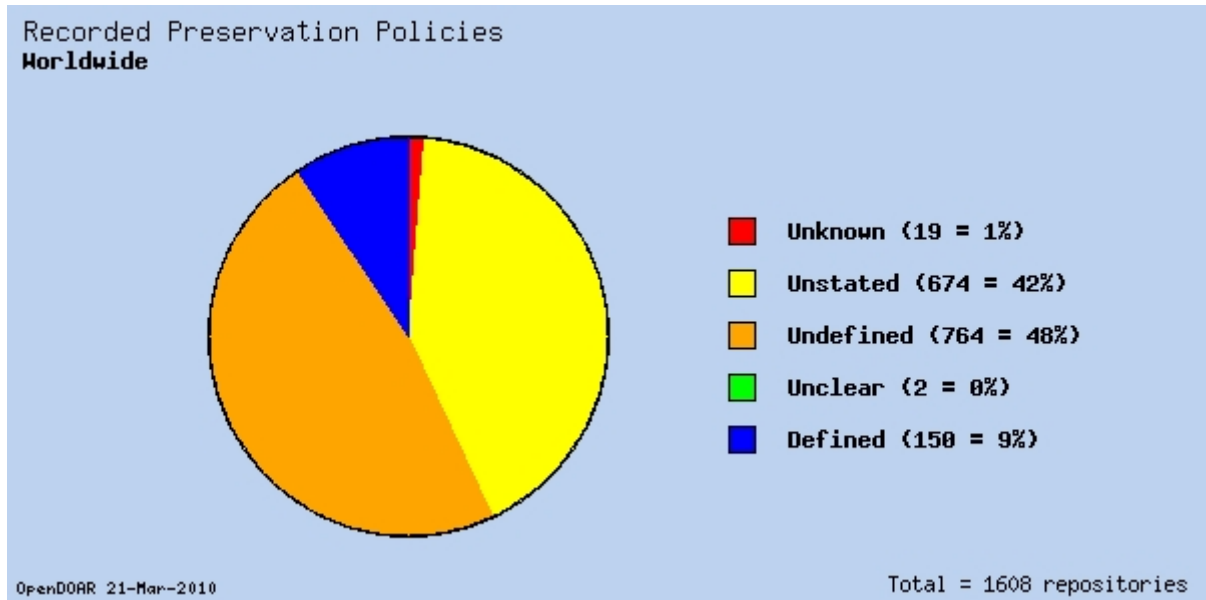
Bisogna evidenziare che in Italia, sebbene la preservazione sia la missione istituzionale dei depositi, attualmente questi sembra che si concentrino sulla selezione e sull'acquisizione di una massa critica di contenuti digitali, oltre che di promuovere una cultura di immissione, raccolta e condivisione dei dati da parte degli autori, e di fornire l'accesso al materiale raccolto al momento. I metadati e gli aspetti tecnici richiesti per la preservazione dei metadati e degli identificatori non sono ancora ben organizzati nei depositi delle istituzioni universitarie (Sacchi 2005; Tammara 2007; Bevilacqua 2008; Cassella 2008).

2.2 Metodologie per la preservazione

Il modello di riferimento standard per la preservazione dei depositi digitali è il modello OAIS, che è stato sviluppato con lo scopo di garantire l'accesso nel tempo alle risorse digitali. Molti aspetti del modello OAIS sono applicabili ad ogni tipo di deposito istituzionale. Attualmente ci sono quattro diversi approcci per eseguire le attività necessarie alla preservazione:

² Accessibile a:<hdl.handle.net>

migrazione, preservazione della tecnologia, emulazione, preservazione dell'oggetto persistente (Smith 2003). Il modello di riferimento OAIS, che è diventato un Information Standard Organization (ISO) nel giugno 2003, è stato adottato ampiamente dai depositi istituzionali.



La figura 2 tuttavia evidenzia che solo il 9% dei depositi istituzionali ha una politica definite per la preservazione. Questo può essere in parte motivato dal fatto che, esclusi i depositi istituzionali con una gran quantità di documenti, che sono anche quelli che esistono da decine di anni, tutti gli altri depositi in Italia sono relativamente recenti e non hanno ancora affrontato cambiamenti tecnologici complessi su larga scala. Questo significa che la strategia più diffusa di preservazione dei depositi istituzionali, tra le quattro definite sopra, è la migrazione, ad esempio da una versione di software ad un'altra: Inoltre questo evidenzia la prevalenza attuale di software commerciale per la produzione di contenuti digitali, come Microsoft Office e Adobe. Mentre una politica per la preservazione ben definita, dovrebbe concentrarsi su software obsoleti, sulla migrazione dell'hardware e il "refreshing" dei media, questi problemi non vengono considerati sufficientemente dai depositi istituzionali, che sono attualmente soprattutto alle prese con la raccolta e l'organizzazione del flusso corrente dei contenuti digitali.

Tutti i depositi istituzionali usano o dovrebbero usare la terminologia e il modello concettuale degli archivi OAIS. Tuttavia, questo uso non è ancora così prevalente, anche perché i depositi istituzionali sono nati in organizzazioni universitarie e di ricerca che avevano già i loro sistemi informativi, i loro fornitori e processi interni di un certo tipo. Alcuni sforzi sono ora in corso da parte dei depositi istituzionali di immettere dati di archivio (ingestion) con il Submission Information Package (SIP) e di produrre dei pacchetti Dissemination Information

Packages (DIP), per rispondere allo spirito della norma ISO. I depositi istituzionali stanno facendo questa attività di ridisegnare le funzionalità di base per la necessità di aumentare l'interoperabilità, ed è probabile che il modello OAIS diventerà ancora più diffuso, come la base concettuale standard per i depositi istituzionali scientifici.

3. Infrastruttura nazionale

Tra le buone pratiche dei depositi istituzionali va considerato il lavoro fatto da OCLC ed il Research Libraries Group nel 2002 per definire gli attributi essenziali che vanno considerati per i depositi istituzionali nell'ambito delle organizzazioni di ricerca "for trusted, reliable, sustainable digital repositories" (Research libraries group 2003).

Vista la complessità dei problemi di preservazione, chi ha la responsabilità di questa non può essere un singolo (ad esempio un privato, come un editore, o anche un singolo creatore di contenuti che apre un suo deposito perché non darebbe sufficienti garanzie) ma deve essere un'istituzione pubblica con una prospettiva di vita permanente. Ma la preservazione di lungo periodo non può essere solo responsabilità di un'università. Un'infrastruttura per le pubblicazioni digitali deve essere il risultato di un approccio in molte direzioni, che coinvolge gli editori scientifici, le biblioteche, gli archivi, le istituzioni di ricerca e tutti gli interessati, inclusi gli utenti ed i creatori di contenuti durante il ciclo di vita delle pubblicazioni digitali. La preservazione delle pubblicazioni scientifiche digitali pone sfide specifiche al riguardo al governo, che deve dare un supporto appropriato, insieme a fondazioni ed altre fonti di finanziamento. Il problema delle responsabilità è collegato a quello dei finanziamenti, come anche a quello dei possibili risparmi che si possono realizzare con la cooperazione. La preservazione è una delle attività dei depositi istituzionali per cui la cooperazione ha un'importanza chiave, ma le problematiche economiche restano.

3.1 Costi e modello economico della preservazione

Un problema che spesso viene trascurato in tutte le sue implicazioni dagli sviluppatori di depositi istituzionali riguarda la necessità di un'attenta programmazione dei costi e soprattutto di chi li dovrà sostenere. C'è una crescente quantità di letteratura sui modelli di costo dei depositi istituzionali, in particolare utili guide sono recuperabili nei seminari Leadirs (LEADIRS, 2004). Tuttavia, i costi della preservazione non sono ben noti. Modelli economici che assicurino la sostenibilità dei programmi di preservazione sono però tuttora poco chiari.

Le università ed istituti di ricerca che avviano un deposito istituzionale possono decidere di trovare il finanziamento necessario dalle proprie risorse interne oppure da finanziamenti esterni. La raccolta di informazioni sui costi di archiviazione e conservazione digitale si è rivelato difficile nella indagine fatta dalla Fondazione Rinascimento Digitale, con la maggior

parte di coloro che hanno risposto che non sono in grado di fornire informazioni sui costi. Il modello economico sembra questo “il finanziamento ora c'è, e si fa finché c'è”.

3.2 Deposito legale del digitale

Lo scopo del deposito legale è quello di assicurare l'accesso a tutte le pubblicazioni di una nazione per il lungo periodo da parte della biblioteca nazionale o da un insieme di biblioteche definite. Per attuarlo, ci deve essere una legge o un regolamento che deve obbligare tutti i creatori di contenuti a mandare copie della pubblicazione alla biblioteca nazionale e/o ad altre biblioteche. In ambito digitale, questa legge non c'è, per alcune difficoltà che in parte sono già state elencate: la natura flessibile della pubblicazione digitale, la mancanza di cataloghi e bibliografie che descrivano le pubblicazioni digitali, la difficoltà di un controllo bibliografico dell'editoria digitale, la difficoltà di individuare il creatore di contenuti. Tutte queste problematiche hanno reso la stesura di una legge sul deposito delle pubblicazioni digitali lenta e incerta. L'ostacolo maggiore ad una legge sul deposito legale è stato però rappresentato dal copyright: poiché la copia digitale di una pubblicazione può essere facilmente essere resa accessibile, copiata e distribuita, gli editori chiedono che la protezione dei diritti di proprietà sia assicurata in qualche modo.

3.3 Sistemi cooperativi di preservazione

Ci sono degli interessanti esempi di collaborazione per la preservazione che stanno nascendo, come un possibile servizio del consorzio CILEA. Fuori d'Italia, alcune grandi organizzazioni (ad esempio OCLC, DSpace, National Library of Australia) stanno cercando di offrire servizi a valore aggiunto ai depositi istituzionali, come modi di razionalizzare i costi delle attività di conservazione.

L'avvento di soluzioni “off-the-shelf” affidate all'esterno mostra che si sta arrivando alla maturità nel concetto della preservazione del digitale. I sei sistemi più diffusi, OCLC Digital Archive, DSpace, LOCKSS (Lots of Copies Keep Stuff Safe), Fedora™, PANDAS, ed il Digital Information Archive System (DIAS) di IBM, provengono da diversi tipi di istituzioni: un fornitore mondiale di servizi per biblioteche, un deposito istituzionale universitario, una biblioteca di ricerca universitaria e un fornitore di servizi editoriali, un deposito universitario insieme ad un gruppo di ricerca di una biblioteca digitale, un sistema di biblioteca nazionale e una biblioteca nazionale. Queste esperienze di istituzioni diverse che collaborano insieme, con istituzioni pubbliche che collaborano con una compagnia commerciale, dimostrano la necessità di partnerships e di interazione fra vari soggetti del settore.

La cooperazione e le nuove partnership sono sempre più importanti. Negli ultimi dieci anni ci si è resi conto che la cooperazione è l'unico modo per assicurare che le pubblicazioni scientifiche digitali siano conservate per il lungo periodo. La cooperazione è stata

inizialmente usate dalle biblioteche per mettere a disposizione una maggiore quantità di pubblicazioni. In aggiunta a questo, nei depositi istituzionali la cooperazione dà la possibilità di assicurare una qualche forma di ridondanza, insieme alla sostenibilità economica dei depositi, con la condivisione dei costi dal punto di vista economico, oltre che i risparmi derivati dal ri-uso dei singoli documenti.

La preservazione e l'accesso a lungo termine rendono la cooperazione necessaria, soprattutto per le istituzioni con meno risorse. Alcuni modelli cooperativi di preservazione digitale che possono essere indicati per i depositi istituzionali sono: Portico, Lockss, Sherpa.

Portico è stato lanciato da JSTOR ed è un'organizzazione no-profit che si offre come archivio di deposito permanente per i periodici elettronici scientifici. Il servizio è sostenuto dalle quote degli editori e dagli abbonamenti delle biblioteche. Gli editori effettuano un deposito volontario dei contenuti dei periodici; il servizio che Portico offre è quello di uniformare i diversi formati e memorizzare i contenuti. Il deposito è centralizzato, con dei mirror distribuiti. Sono previste delle limitazioni ai contenuti accettati per la preservazione: vengono preservati i contenuti degli articoli, compreso eventuali tabelle, immagini ed appendici, ma non le funzionalità delle interfacce o il layout dei periodici. L'accesso è consentito solo alle biblioteche che sono abbonate e non agli utenti finali: è possibile chiedere di avere accesso ad una risorsa solo nel caso in cui l'accesso non sia possibile in altro modo, per circostanze eccezionali, come nel caso che gli editori abbiano chiuso la loro attività, oppure deciso di togliere dal commercio i periodici o per altre circostanze catastrofiche.

Lockss (Giuli, D. et al. 2005) è un progetto basato su una rete distribuita di sistemi "peer to peer" che si propone per la preservazione di periodici scientifici e altri contenuti digitali. Il progetto, coordinato dalla Stanford University, prevede un elevato numero di depositi finalizzati alla preservazione dei contenuti e distribuiti nel Web (Web chaches) indipendenti e con costo limitato ma capaci di assicurare la preservazione in modo cooperativo e proteggere i contenuti da eventuali danni.

Sherpa Digital Preservation è anch'essa un'organizzazione cooperativa, in cui i depositi istituzionali insieme agli archivi istituzionali, nella cornice operativa dello standard OAIS, affidano ad un'istituzione esterna, che in questo caso è Arts and Humanities Data Service AHDS, il compito di assicurare la preservazione di lungo periodo dei contenuti digitali. Questo modello è particolarmente conveniente per le istituzioni più piccole, che non hanno le risorse e le competenze per programmi di preservazione di grande impegno. Un modello di

licenza è stato predisposto per le istituzioni cooperanti, su cui queste potranno basare le politiche interne di immissione dei contenuti.

4. Conclusioni

L'analisi degli elementi necessari per l'accesso permanente alle pubblicazioni scientifiche digitali, sopra elencati esemplificando le esperienze di depositi istituzionali in Italia ed all'estero, può forse aver chiarito il quadro delle problematiche per la preservazione. L'infrastruttura necessaria per garantire l'accesso alle pubblicazioni scientifiche nel 2060, non è ancora stata realizzata, anche se i depositi istituzionali ed il deposito legale volontario delle pubblicazioni digitali possono essere considerati una prima importante realizzazione verso la creazione di un'infrastruttura nazionale. E' tuttavia disponibile il servizio di accesso alle pubblicazioni scientifiche dei depositi istituzionali: questo è assicurato dai registri, che sono dei servizi di natura globale che raccolgono i metadati delle pubblicazioni, che sono necessari per mantenere una descrizione catalografica autorevole, con informazioni sui set dei metadati, informazioni di riferimento per i formati e dati per l'hardware ed il software necessario per l'accesso.

La preservazione dei depositi istituzionali è un'area di ricerca e di applicazione con un futuro incerto. E' prevedibile che non tutti i depositi istituzionali potranno affrontare da soli le problematiche della preservazione a lungo termine, né affrontarne i costi, ancora in gran parte poco noti. Il problema della preservazione non è un problema secondario, ma è legato all'esistenza stessa dei depositi istituzionali, che senza questa garanzia, non potranno convincere gli autori ad immettere i loro contenuti digitali e di conseguenza non potranno realizzare il duplice scopo dei depositi istituzionali.

Ci sono ancora molte problematiche irrisolte per la preservazione e l'accesso permanente alle pubblicazioni scientifiche nei depositi istituzionali, che riguardano ulteriori aspetti legislativi e politici, che richiedono ulteriore discussione da parte di tutte le parti interessate. Per esempio, devono essere risolti gli aspetti politici e legali di creare un deposito federato oppure di sviluppare delle partnership che possano funzionare tra vari interessati e per diverse tipologie di documenti (banche dati, pubblicazioni, multimedia, etc.). Ci sono anche alcuni aspetti tecnici che richiedono ulteriore ricerca, come la generazione automatica di metadati, o la fornitura di meccanismi di archiviazione in strumenti di creazione di contenuti (authoring). Altre questioni tecniche comprendono la creazione di archivi interoperabili e buone pratiche per l'archiviazione e la conservazione del deposito stesso. La ricerca sull'

archiviazione e la preservazione dinamica, di contenuti digitali non-HTML e basati su un database di contenuti Web è anche un'attività di ricerca importante.

In conclusione, qualcosa è stato realizzato di quell'infrastruttura che dovrà consentire ad un ricercatore nel 2060 di poter identificare, localizzare ed usare una pubblicazione scientifica in un deposito istituzionale, ma molto resta ancora da fare per dare la certezza a questo ricercatore di ottenere quello che cerca. Per superare tutte le difficoltà elencate, economiche, giuridiche (come per il copyright) ed anche tecniche, non c'è altra soluzione che questa: occorre sviluppare all'interno della comunità scientifica e della società in generale una cultura che incoraggi la "volontà di preservare e garantire l'accesso permanente".

RIFERIMENTI BIBLIOGRAFICI:

- Bevilacqua, F. (2008). "L'organizzazione dei depositi istituzionali DSpace in Italia." Biblioteche oggi **26**(6): 17-25.
- Branin, J. (2004). Institutional repositories. Encyclopedia of Library and Information Science, Dekker.
- Cassella, M. (2008). "L'Open Access nelle scienze umane." Biblioteche oggi **26**(10): 40-49.
- Giuli, T., R. D., et al. (2005). "The Lockss peer to peer digital preservation system." ACM transactions on Information Systems **23**(1).
- Lynch, C. (1998). "Identifiers and their role in networked information applications." CAUSE/EFFECTS(Winter): p. 8-14.
- Lynch, C. A. (2003). "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age." ARL Bimonthly Report **226**(February).
- Research libraries group (2003). "Preserving our digital heritage." RLG diginews **56**.
- Sacchi, S. (2005). "L'open access negli atenei italiani." Biblioteche oggi.
- Smith, A. (2003). New model scholarship: how will it survive? Washington, CLIR.
- Tamaro, A. M. (2007). "I depositi istituzionali ed i nuovi paradigma della comunicazione scientifica." < <http://dspace-unipr.cilea.it/handle/1889/543>>
- Tamaro, A. M. (2008). Biblioteche digitali e scienze umane. Open Access e depositi istituzionali. Fiesole, Casalini Libri.
- Vitiello, G. (2004). "L'identificazione degli identificatori." Biblioteche oggi **22**(2): 67-80.