

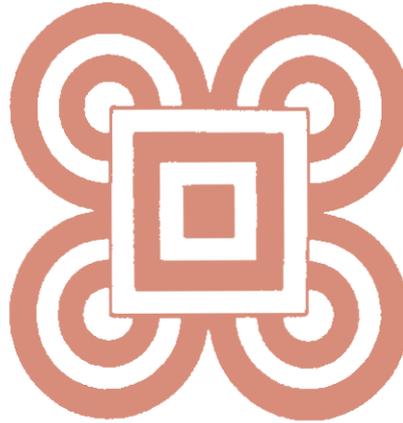
Dspace e i suoi servizi

Un innovativo sistema di biblioteca digitale

Pietro Gozetti

Brescia
pgboy79@tin.it

Un deposito istituzionale è un insieme di servizi che un'università offre alla comunità dei propri membri per la conservazione, la preservazione a lungo termine, l'organizzazione, la gestione dell'accesso e la disseminazione di risorse digitali create dall'istituzione e dai membri stessi.¹ Si tratta di un'attività complessa, che coinvolge diversi profili professionali, dal bibliotecario all'informatico, al *records manager*, agli amministratori e responsabili della struttura universitaria nonché i docenti e ricercatori, come attori e principali fruitori dei servizi. Un deposito digitale è un nuovo canale di comunicazione scientifica veloce ma al contempo autorevole, coerente con l'economia del mezzo elettronico, che garantisce accesso a quelle risorse in genere confinate nei pc, nei siti web e nei server dipartimentali, e che garantisce stabilità nel tempo e leggibilità dei formati; per il momento rappresenta una soluzione complementare, piuttosto che sostitutiva, rispetto ai periodici cartacei, ma la crescita di questo modello procede a ritmi vertiginosi. Gli editori di riviste scientifiche, nel frattempo, sostengono che sia improbabile che tali depositi soppiantino le loro pubblicazioni, ritenendo che i periodici siano ancora il mezzo migliore per diffondere e salvaguardare la ricerca; anche alcuni tra i sostenitori dei depositi istituzionali riconoscono la difficoltà di persuadere certi autori a cambiare le proprie abitudini di pubblicazione. In diverse facoltà sono nati server di e-print e pre-print, grazie al fat-



to che i depositi istituzionali facilitano l'accesso a documentazione difficilmente consultabile, o perché antica (digitalizzata) o perché di difficile reperimento (letteratura grigia); inoltre, nei casi di discipline più restie al cambiamento, i depositi digitali hanno la funzione di aiutare le facoltà a farsi addirittura iniziatrici di questo nuovo tipo di trattamento della produzione scientifica. L'attività di conservazione e diffusione della produzione scientifica di un ateneo, sia per i costi che per un'organizzazione efficace ed economica, va progettata secondo tecniche manageriali: deve essere approvata dall'alto e venire attuata basandosi su riscontri statistici relativi all'attività svolta dai membri della comunità, alle finalità, ai bisogni, alle aspettative d'uso del servizio. Inoltre, una volta implementato il deposito, deve essere possibile verificare il raggiungimento degli scopi e, se necessario, migliorare o rettificare il servizio. Un deposito digitale non è infatti un modello statico, ma si adatta e sussiste in funzione della struttura organizzativa dell'ente produttore

e delle sue finalità. Ovviamente il sistema dovrà basilarmente contenere i sussidi generati da insegnanti e studenti (tesi, e-print, statistiche) ma anche documentazione relativa alla vita culturale dell'ateneo, rapporti, eventi, simposi.²

1. Cos'è DSpace?

DSpace è un innovativo sistema di biblioteca digitale per catturare, immagazzinare, indicizzare, preservare e ridistribuire la produzione intellettuale in formato digitale di una facoltà universitaria di ricerca, in altre parole un deposito istituzionale di facile utilizzo, grazie a un'interfaccia utente basata sul web, che può essere adattato a istituzioni e singoli dipartimenti. DSpace trova la sua origine in eventi importanti per l'approccio open access, come il convegno "Universal pre-print server" (1999),³ organizzato da diverse organizzazioni bibliotecarie e dalla Research Library del Los Alamos National Laboratory, la Open Archives Initiative,⁴ la Budapest Open Access Initiative⁵ e la Convenzione di Santa FE.⁶ Nel 2000 il Massachusetts Institute of Technology (MIT), gli Hewlett Packard Labs e un comitato di consulenza composto da ricercatori delle facoltà interessate hanno iniziato a lavorare sul software DSpace grazie a un finanziamento di 1,8 milioni di dollari.⁷ Nel novembre del 2002 il sistema è stato reso disponibile open source alle istituzioni di ricerca di tutto il mondo, e tra la primavera e l'estate 2002 si è svolta la fase di prova, du-

rante la quale DSpace è stato adottato da quattro istituzioni di ricerca: la Sloan School of Management, il Department of Ocean Engineering, il Centre for Technology, Policy and Industrial Development e il Laboratory for Information and Decision Systems.⁸ Gli scopi di questo periodo sperimentale erano:

- 1) ottenere consapevolezza di come i collaboratori del MIT avrebbero usato il sistema;
- 2) informare riguardo le caratteristiche che gli utenti avrebbero trovato utili o avrebbero preferito fossero meglio sviluppate;
- 3) prevedere i costi al fine di tracciare un piano finanziario a lungo termine.

La crescita e la forma di DSpace dipendono da ciò che gli utenti vogliono veramente, in base ai dettagli della recente prospettiva utente-centrica: se fino a qualche anno fa i bibliotecari si rivolgevano al cosiddetto "utente medio", indistinto e anonimo, ora i servizi tendono a complicarsi e a diversificarsi in virtù delle esigenze della varietà di utenti che frequentano biblioteche, banche dati e risorse su web.

1.1. Potenzialità e tipologie di contenuti

DSpace fornisce un modo economico per gestire materiali e pubblicazioni a scopo di ricerca entro un deposito che viene mantenuto in maniera professionale, allo scopo di dar loro una maggiore visibilità e accessibilità nel tempo, un contenitore che ha anche il grande vantaggio di essere personalizzabile secondo le esigenze delle singole istituzioni. Le informazioni depositate in DSpace possono essere accessibili con qualsiasi motore di ricerca, e gli studiosi hanno facoltà di scegliere se usarlo come deposito temporaneo prima della pubblicazione definitiva (preprint), o come mezzo per archiviare stabilmente la loro produzione. Ogni

facoltà sceglie cosa inserire nel sistema e chi è autorizzato a farlo, e in genere i depositi sono organizzati per comunità (una scuola, un dipartimento, un laboratorio ecc.); il software accetta tutti i formati di file: testo, immagini, video, audio, articoli e preprint, rapporti tecnici, working paper, tesi elettroniche, dataset, learning object, collezioni di documenti digitalizzati, letteratura grigia.

Nell'ambito di questo sistema possono essere utilizzate proficuamente le professionalità già esistenti: ad esempio i catalogatori possono ricoprire un ruolo appropriato nel precisare i metadati o gestire il database di DSpace; i tradizionali reference librarian possono analogamente prestare lo stesso aiuto ai fruitori del deposito digitale.

1.2. La tecnologia di DSpace

DSpace è disponibile dietro licenza *BSD open source*, scaricandolo da SourceForge, per essere utilizzato così com'è, oppure per essere modificato ed esteso secondo le necessità; è programmato in Java e funziona su qualsiasi sistema operativo UNIX o LINUX.⁹ È inoltre possibile reperire documentazione tecnica e sull'installazione e unirsi alla *DSpace tech-mailing list*, formata da altri sviluppatori di DSpace, nonché usufruire del supporto in linea tramite FAQ. Trattandosi di un sistema aperto e gratuito, il suo codice è disponibile agli sviluppatori per essere esteso o corretto nel caso si riscontrino dei bug. Robert Tansley, sviluppatore di punta del team della HP, ha diviso il software in moduli interscambiabili, sostituibili via via che nuove versioni si rendono disponibili, rendendo meno dispendioso l'upgrade.¹⁰ Il codice d'origine è programmato in Java, mentre gli altri includono un sistema di gestione del database relazionale (PostgreSQL) e un server

http Apache. L'assetto tecnologico di DSpace consente piena interoperabilità con altri archivi digitali, favorendo quello che Besser ha definito il passaggio da collezioni digitali isolate a quelle intercomunicanti.¹¹

1.3. La DSpace Federation

La Federazione di DSpace (DSF)¹² è un'organizzazione informale, creata grazie al contributo della Andrew W. Mellon Foundation e composta da atenei e istituzioni di ricerca tra le quali le più importanti sono: Columbia University, Cornell University, Ohio State University, University of Rochester, University of Toronto, University of Washington a Seattle, Cambridge University (unico ateneo inglese), Massachusetts Institute of Technology (MIT).

I suoi membri perseguono i seguenti scopi:

- promuovere l'interoperabilità tra i depositi istituzionali per sostenere servizi distribuiti, comunità e collezioni virtuali;
- collaborare per lo sviluppo e il mantenimento del codice di DSpace;
- sviluppare un corpus critico di contenuti che rappresentino la produzione intellettuale delle istituzioni di ricerca più importanti nel mondo;
- promuovere lo sviluppo continuo del servizio DSpace;
- assicurare la conservazione a lungo termine della produzione scientifica compatibilmente con diversi standard, e sostenendo iniziative nazionali e internazionali per implementare nuovi standard;
- offrire documentazione tecnica e assistenza per programmare e costruire DSpace.

2. Pianificare, implementare e definire il proprio servizio DSpace

Alla base del successo dei depositi istituzionali esiste sempre la pro-

gettazione e la pianificazione: sebbene l'installazione e il settaggio siano facili, progettare e costruire il proprio DSpace impone scelte strategiche determinanti tra cui come la propria università userà DSpace, quali tipi di contenuti vi immagazzinerà e chi potrà contribuirvi.

Una volta avviato il progetto DSpace all'interno di un'istituzione di ricerca, è necessario produrre coscienza nella medesima e diffondere l'idea del nuovo servizio tra lo staff e tra gli utenti finali, testandolo entro un gruppo o comunità nell'ateneo e fornendo contestualmente il supporto costante da parte dello staff bibliotecario e tecnico-informatico.

2.1. Le sfide da affrontare

Proporre un superarchivio della portata di DSpace implica indubbiamente per gli sviluppatori sia delle scelte politiche e strategiche sia delle questioni finanziarie. La problematica iniziale da affrontare è innanzitutto quali sono i destinatari. Porsi questa domanda non è inutile: c'è molta differenza tra progettare un servizio destinato a un'utenza ristretta e precisa, e pianificare invece un sistema che debba essere destinato a un utilizzo generale e indefinito.

DSpace è simile a una scatola vuota, il cui valore è proporzionale al contenuto:¹³ i servizi che offre sono veramente significativi se creano una rete di fruizione abbastanza ampia da giustificare gli sforzi intrapresi. Quindi il vero interrogativo sul futuro di DSpace si sposta sugli attori principali, ovvero i creatori-fruitori, che devono essere i primi a curare l'elaborazione e la gestione delle loro opere, rompendo la catena che (per una prassi editoriale apparentemente ormai ineludibile, o per negligenza) priva del controllo totale sulle proprie pubblicazioni.

Ci sono campi di studio che dimostrano maggiore sensibilità riguardo all'archiviazione autogestita e che sono già attivi da tempo in tal senso, si pensi alla fisica o all'elettronica. La prima iniziativa per sviluppare un sistema centralizzato di disseminazione dei preprint attraverso gli Information Exchange Groups nell'ambito delle scienze biomediche è del 1961, mentre sempre negli anni Sessanta lo Stanford Linear Accelerator Center (SLAC) rivestì un ruolo guida nell'iniziativa di raccolta e catalogazione di preprint nel campo della fisica delle alte energie (HEP), sostenendo quindi una cultura dei preprint in quella disciplina; fu proprio in questo campo che si venne sviluppando il primo server di preprint ai Los Alamos National Labs nei primi anni Novanta.¹⁴

L'auspicio ora è che anche altre discipline (soprattutto umanistiche) prendano coscienza dell'urgenza di una tale soluzione collaborativa, abbandonando un sistema di lavoro che troppo spesso privilegia il singolo o il piccolo gruppo e lasciando da parte le reticenze o le negligenze verso la cultura in formato digitale (quanti studiosi ancora stampano le loro bozze per correggerle a mano e poi ricopiarle?). Il più grande ostacolo è perciò l'inerzia: i professori non useranno un deposito digitale se lo percepiscono come un lavoro in più, anche se in linea di principio ne apprezzano l'idea.

2.2. La questione della peer-review e delle abitudini della comunità scientifica: un vicolo (quasi) cieco

Il sistema di comunicazione sotteso da DSpace implica delle innegabili ripercussioni nei processi di critica, citazione e recensione che vanno sotto il nome di peer-review: nell'ambito dei depositi istituzionali la necessità più impellente è quella dell'accesso, ma la

qualità delle risorse a carattere scientifico deve essere comunque garantita. È palese la necessità di applicare nuovi modi di certificazione del testo: secondo una ricerca effettuata da Steve Lawrence, un ricercatore in scienze informatiche del NEC Research Institute, gli articoli che appaiono online hanno una probabilità del 7,03 per cento di essere citati da altri ricercatori rispetto al 2,74 per cento di quelli che non vengono pubblicati online.¹⁵ È indubbio quindi che gli studiosi non diffidano del materiale online, specialmente se è disponibile gratuitamente in depositi istituzionali e prodotto da altri studiosi da loro conosciuti; anzi, molti ricercatori hanno affermato che il feedback che ottengono dalla disseminazione su un server di preprint è più prezioso dei commenti redatti da un ristretto numero di critici e forniti dagli editori. Se la qualità delle pubblicazioni proposte fosse scadente, quale professore serio le citerebbe? Questo è vero, ma significherebbe anche ridurre un processo intellettuale come la peer-review a un mero calcolo citazionale, e si è addirittura ipotizzato di sostituire la peer-review con altri sistemi come il *ranking*, che non sono ancora tanto perfezionati da poter competere con essa (se mai possono esserne un supplemento).¹⁶ La peer-review non è un'attività obsoleta e da abbandonare in quanto implica consumo di tempo e denaro: in un deposito digitale possono benissimo convivere documenti transitori (letteratura grigia, programmi dei corsi, avvisi ecc.) e documenti sottoposti a peer-review, destinati alla conservazione permanente. La certificazione della qualità degli articoli, che oggi deve basarsi su standard ben definiti e quantificabili, nonché dare la possibilità ai lettori di interagire commentando gli articoli (*peer commentary*), è possibile:

il problema sta nell'attribuzione di questa competenza, dal momento che non esiste più un modello centralizzato convergente nell'editore il quale coordinava la revisione. Nel nuovo modello distribuito sembra più facilmente proponibile che le commissioni di pari vengano contattate direttamente dall'autore o dall'istituzione cui fa capo il deposito istituzionale, prima dell'inserimento dell'articolo nel deposito stesso. L'istituzione è l'organismo che garantisce la qualità delle pubblicazioni ospitate sui propri server e depositi, e deve perciò definire un workflow uniforme (vedi oltre), valutare la coerenza delle risorse ospitate con la politica di sviluppo dei depositi, verificare l'esattezza della descrizione dei documenti e la loro rispondenza a parametri e standard largamente condivisi.

2.3. Chi deve inserire i documenti in DSpace?

Altro fattore di incertezza consiste nel decidere quali siano i soggetti autorizzati a pubblicare o a rimuovere i documenti dal deposito. La strada più facilmente praticabile è quella di lasciare che ogni singola comunità di studiosi decida chi ha l'autorizzazione a caricare o rimuovere i documenti.

È a discrezione delle singole comunità anche la decisione sulle modalità di permanenza dei documenti nel deposito. Una parte di ciò che viene ospitato nei depositi è materiale di pre-pubblicazione, destinato all'edizione su periodico: a seconda degli accordi presi con l'editore, il preprint dovrà essere rimosso op-

pure potrà rimanere ospitato permanentemente nel deposito istituzionale. Resta sempre ai responsabili delle comunità il compito di decidere la rimozione o la modificazione dei documenti una volta inseriti e il destino dei materiali ospitati da un deposito, quando l'istituzione cui afferisce chiude o si trasforma. Come norma generale sarebbe utile che, una volta inseriti dei documenti in un deposito, a meno che non violino le leggi sul copyright, essi vi restassero ospitati in maniera definitiva, in modo da garantire ai depositi stessi la maggiore stabilità e affidabilità possibile.

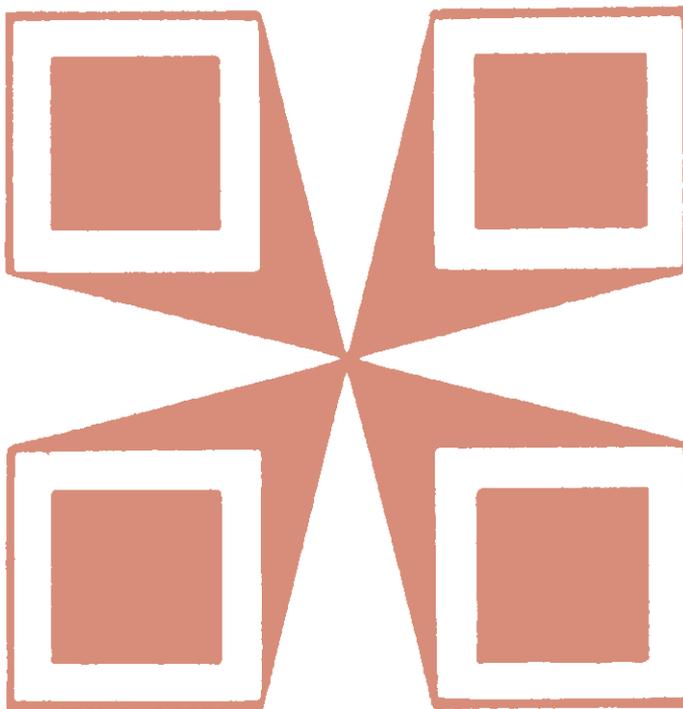
2.4. Cosa si deve inserire in DSpace?

Per quanto riguarda la tipologia dei materiali ospitati nel deposito, anche qui non sono poche le discussioni fra coloro che vorrebbero dei criteri selettivi in base ai formati, ai contenuti o alle finalità dei documenti e coloro i quali invece desidererebbero un'ampia copertura da parte del deposito, col rischio però di sovraccaricarlo. Un de-

posito istituzionale consente di mettere a disposizione di un vasto pubblico anche tipologie di dati che in ambito cartaceo non avrebbero potuto circolare (ad esempio i documenti audiovisivi), o che sarebbe stato troppo costoso stampare (come le fotografie). Analizzare le funzioni e le attività dell'ente presso il quale è implementato il deposito, e in base a ciò decidere quale materiale conservare e come immagazzinarlo, è una decisione politica necessaria alla crescita corretta e coerente del sistema. È consigliabile inoltre che i file siano depositati in formati ampiamente leggibile e comunque coerenti.

2.5. I metadati

I documenti inseriti in DSpace sono corredati da una serie di metadati che permettono il recupero del documento anche da un comune motore di ricerca, con una priorità (*ranking*) elevata. DSpace sfrutta l'Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH),¹⁷ che permette ai metadati di venire indicizzati insieme ad altri metadati per essere reperiti tramite un'unica ricerca tra i documenti relativi a un argomento particolare. Il Libraries Working Group Application Profile è lo standard per la descrizione, basato su Dublin Core: solo tre campi sono obbligatori (titolo, lingua, data di accesso nell'archivio), mentre tutti gli altri sono opzionali (abstract, parole chiave, informazioni tecniche e sui diritti di proprietà).¹⁸ In fase di ricerca verranno visualizzati, nel record risultante, tutti i metadati presenti relativi a quel documento, grazie a uno schema basato su XML.



2.6. Workflow

Per poter inserire materiali in una collezione del deposito è necessario anzitutto disporre di un'autorizzazione da parte degli amministratori del sistema e quindi entrare a far parte di una comunità. La submission avviene per gradi, durante i quali il documento non è ancora accessibile al pubblico, ma solo agli attori del workflow: in primo luogo l'utente deve inserire i metadati riempiendo una maschera-guida di riferimento, in seguito dovrà inviare il documento e visionare l'insieme di diritti e licenze di accesso che desidera applicare ai propri materiali. Tutte queste operazioni vengono rese il più possibile semplici anche per l'utente più inesperto, grazie all'help contestuale disponibile per ogni passo della submission.

A questo punto il materiale viene sottoposto a un processo di validazione atto a verificare la rispondenza del medesimo a determinati standard di qualità. La risorsa viene sottoposta a peer-review (valutazione del contenuto e della veste editoriale) e può anche venire rifiutata e rispedita all'autore corredata dalle motivazioni del rifiuto e da consigli per le modifiche da apportare. A partire da questo punto, se il documento viene accettato non può più essere rifiutato, ma solo retrocesso alla fase precedente. La validazione prosegue con la verifica della congruità dei metadati utilizzati dall'autore (e quindi si provvede all'integrazione e/o modifica in caso di mancanze o errori) e con un'ultima revisione formale prima della immissione vera e propria. A ogni passaggio da uno stadio al successivo, una e-mail inviata dal sistema avvertirà gli attori del workflow circa la presenza di un nuovo documento da esaminare.

3. I problemi aperti dei depositi istituzionali

Gli ostacoli e i problemi aperti dei depositi istituzionali riguardano questioni comuni all'accesso e all'organizzazione stabile e coerente di oggetti e collezioni digitali, quali il copyright, la preservazione e la corretta identificazione. Si elencano di seguito le soluzioni avviate da Dspace, ancora come scelte iniziali da consolidare con sforzi organizzativi anche interistituzionali.

3.1. La questione del copyright in Italia

In ambito digitale la legge ha gli stessi effetti che nell'ambito cartaceo, stando alla dottrina, dal momento che non esistono limitazioni all'art. 1 della legge n. 633 del 22 aprile 1941.¹⁹ La legge sembra però non prevedere atti scritti per la pubblicazione dei testi digitali: la cessione dei diritti economici avviene tramite contratto nel caso di edizione a stampa,²⁰ mentre il testo non fa riferimento a documenti digitali. Purtroppo le normative e la sensibilità dei legislatori non hanno tempestivamente seguito il cambiamento dei canali di comunicazione scientifica. L'ultimo aggiornamento alla suddetta legge, il d.lgs. n. 68 del 9 aprile 2003, non risolve una questione centrale come la riproduzione dei testi digitali. Si dispone comunque che la riproduzione sia lecita presso i terminali delle istituzioni di ricerca, biblioteche e musei, e quindi subordinata al fatto di trovarsi nei locali dell'istituzione.²¹ Uno dei punti più criticati della normativa sul copyright è poi che essa non fa alcuna distinzione tra canali di ricerca e canali commerciali, mentre in realtà gli autori scientifici non ricevono *royalties* dai loro lavori come i romanzieri, non essendo l'attività scientifica finalizzata al lucro. Se analizziamo la situa-

zione in ambito internazionale si deve rilevare che la transnazionalità delle risorse elettroniche e le divergenze legislative, filosofiche, economiche e sociali rendono difficile l'elaborazione di accordi mondiali sul copyright: molto più economica sembra essere l'elaborazione di diversi tipi di tutela in ambito digitale, a livello contrattuale (licenze elettroniche, disclaimer di siti, condizioni di accesso) e tecnologico (sistemi di protezione contro la riproducibilità). Ormai da diverso tempo infine è invalso l'uso del termine *copyleft*, per designare una prassi pienamente legittima e legale che sta riscuotendo un ottimo successo soprattutto in ambito di circolazione dei software. Il *copyleft* racchiude in sé l'idea di accesso e uso liberi, a patto che la riproduzione e la diffusione dei materiali (ma anche e soprattutto dei prodotti da essi derivati e/o ispirati) non avvengano a titolo oneroso da parte di un distributore diverso dall'autore (o dal detentore per cessione esplicita dei diritti economici). Un incontro tra esigenze di accesso e necessità di protezione del diritto degli autori è costituito da Creative Commons, un'ampia gamma di licenze che spaziano dalla protezione totale del copyright all'assenza di diritti riservati.²² Dspace (nella sua ultima release 1.2) ha recepito tale innovazione, rendendo possibile scegliere una licenza Creative Commons al termine del processo di submission dei materiali.

3.2. La questione della preservazione

Un deposito istituzionale basato sull'open access non può essere considerato in assoluto un archivio a lungo termine, né vi sono iniziative, finalizzate alla preservazione, che possono dirsi concluse. Ciò a causa degli elevati costi associati agli sviluppi tecnologici richiesti e delle difficoltà nell'assegnare re-

sponsabilità a intraprendere tali investimenti, non essendovi un ritorno immediato (solo gli organi politici potrebbero sostenere i costi di tale operazione). La Library of Congress e alcuni grandi editori commerciali hanno potuto contare su finanziamenti cospicui finalizzati alla preservazione nell'ambito dei loro domini, ma è improbabile che i secondi si preoccupino di un futuro molto lontano. La preservazione infatti implica la creazione di formati stabili, o per lo meno l'analisi di quali saranno i formati e gli standard che avranno vita più lunga. Molte università non hanno nemmeno il sostegno finanziario e le competenze per assicurare la preservazione a breve termine, quindi è inevitabile il ricorso a strategie di collaborazione. Nessuna iniziativa organica è stata ancora avviata e i singoli progetti mostrano divergenze sul tipo di approccio da tenere nella risoluzione del problema. I principali approcci sono:

- 1) il mantenimento della stabilità dei supporti d'origine;
- 2) l'emulazione della tecnologia hardware e software;
- 3) la migrazione dei formati dei dati nelle nuove tecnologie man mano che esse si evolvono.²³

Una contemperanza di questi tre metodi è la soluzione su cui posa il consenso generale, anche se i vari progetti danno priorità all'uno o all'altro aspetto.

DSpace utilizza l'OAIS (Open Archival Information System), un modello logico-funzionale che garantisce stabilità e soprattutto interoperabilità, essendo compatibile con altri archivi open access.²⁴ Esso implica una visione dinamica dei metadati destinati alla conservazione, i quali vengono considerati parte integrante di un Archival Information Package (AIP), formato dall'oggetto digitale intrinsecamente legato ai propri metadati. I vantaggi dell'adozione di un mo-

dello uniforme sono intuibili: più agevole interscambio, terminologia comune, maggiore economicità nell'indicizzazione.

3.3. Identificazione, inalterabilità e persistenza delle risorse digitali

Identificare le risorse online contro la loro instabilità è un'operazione necessaria, per quei documenti che si desidera siano sempre accessibili e protetti dalla manipolazione e dal plagio. Tra i migliori codici di identificazione, quello che ha avuto maggiore diffusione è finora il DOI (Digital Object Identifier),²⁵ largamente usato sia per la buona interoperabilità che per la persistenza garantite. L'identificazione nell'ambito dell'OAI, invece, non è legata all'oggetto depositato al suo interno, bensì all'operazione di estrazione dei metadati: un identificatore OAI potrà ospitare degli altri identificatori relativi all'oggetto stesso (ad esempio il DOI).

Identificare una risorsa digitale significa descriverla nei suoi tratti distintivi che la rendono diversa da altre analoghe. Questi tratti distintivi si concretizzano nei seguenti elementi:

- URI (Uniform Resource Identifier) è l'insieme dei nomi e degli indirizzi che si riferiscono alla risorsa;
- URL (Uniform Resource Locator) è il termine che designa la localizzazione della risorsa, che può mutare nel tempo;
- URN (Uniform Resource Name) è l'identificatore del nome della risorsa: può essere sia una URI, che ha l'incarico istituzionale di mantenere stabilità, persistenza e disponibilità all'accesso (e in questo caso può anche essere una URL, come ad esempio la PURL, Persistent Uniform Resource Locator), sia uno schema specifico sganciato da particolari localizzazioni;
- URC (Uniform Resource Characteristic) è un insieme di dati relativi

vi alle caratteristiche specifiche del documento (metadati).

Esiste inoltre una serie di tecnologie di *error detection*, finalizzate alla scoperta di alterazione nei file. Una di esse è il CRC (Cyclic Redundancy Check), in grado di riconoscere gli errori di trasmissione di dati: i messaggi trasmessi sono costituiti da pacchetti di lunghezza predefinita i quali sono divisi da un numero divisore fisso, ancorato ai pacchetti, che viene spedito insieme ad essi; quando il messaggio viene ricevuto il computer ricalcola la dimensione del messaggio in base al valore numerico suddetto e la confronta con esso; ovviamente se i valori non coincidono, viene rilevato un errore. Un'altra soluzione è la *checksum*, un semplice sistema matematico per verificare l'integrità dei dati digitali mediante la somma di controllo dei byte che li compongono (trasmessa assieme ai dati cui si riferisce): anche un singolo bit modificato nei dati comporterà una discrepanza tra la *checksum* di partenza e quella finale.

Bibliografia

- SALLY ATWOOD, *MIT's superarchive*, "Technology Review", December 2002/January 2003, <<http://www.technologyreview.com/articles/atwood1202.asp>>. D.l. 9 aprile 2003 n. 68, *Attuazione della direttiva 2001/29/CE sull'armonizzazione di taluni aspetti del diritto d'autore e dei diritti connessi nella società dell'informazione*, "Gazzetta ufficiale" n. 87 del 14 aprile 2003 – S.O. n. 61.
- ANTONELLA DE ROBBIO, *Metadati per la comunicazione scientifica*, "Biblioteche oggi", 19 (2001), 10, p. 54-60, <<http://www.bibliotecheoggi.it/2001/20011005401.pdf>>.
- Id., *Relazione tecnica su DSpace (MIT)*, E-Lis eprint archive, 23 ottobre 2002, <<http://eprints.rclis.org/archive/00000032/01/DSpace.pdf>>.
- Id., *URI, URN e URL: una questione di de-*

finizioni, "Biblioteche oggi", 20 (2002), 1, p. 30-37, <<http://www.bibliotecheoggi.it/2002/20020103001.pdf>>.

STEVEN HARNAD, *Post-Gutenberg Galaxy: the fourth revolution in the means of production of knowledge*, "Public-Access Computer Systems Review", 2 (1991), 1, p. 39-53, <<http://cogprints.ecs.soton.ac.uk/archive/00001580/00/harnad91.postgutenberg.html>>.

Id., *The PostGutenberg Galaxy: how to get there from here*, "Information Society", 11 (1995), 4, p. 285-292, <<http://cogprints.ecs.soton.ac.uk/archive/00001689/00/thes.html>>.

Id., *For whom the gate tolls? How and why to free refereed research literature online through author/institution self-archiving, now*, (1998), <<http://cogprints.ecs.soton.ac.uk/archive/00001639/01/resolution.htm>>.

Id., *The self-archiving initiative*, "Nature", 410 (2002), 1024-25, <<http://cogprints.ecs.soton.ac.uk/archive/00001639/01/resolution.htm>>.

ANICK JESDANUN, *Digital memory threatened as file formats evolve*, "Houston Chronicle", 16 January 2003, <<http://www.chron.com/cs/CDA/story.hts/tech/1739675>>.

IBIRONKE LAWAL, *Scholarly communication: the use and non-use of e-print archives for the dissemination of scientific information*, "Issues in Science and Technology Librarianship", (2002), <<http://www.istl.org/02-fall/article3.html>>.

Legge 22 aprile 1941 n. 633, *Protezione del diritto d'autore e di altri diritti connessi al suo esercizio*, "Gazzetta ufficiale" n. 166 del 16 luglio 1941.

DEBORAH LENARES, *Faculty use of electronic journals at research institutions*, <<http://www.ala.org/ala/acrl/acrlvents/lenares99.pdf>>.

MARGRET LIPPERT – ERIC CELESTE, *DSPACE: sharing MIT's intellectual output*, "MIT Libraries News", 12 (2000), 2, p. 4, <<http://DSpace.org/news/articles/DSpace-sharing.html>>.

Id., *DSPACE at MIT: what users want*, "SPARC e-News", October-November 2000, <<http://DSpace.org/news/articles/DSpace-what-users.html>>.

CLIFFORD A. LYNCH, *Metadata harvesting and the Open Archives Initiative*, "ARL Bimonthly Report", (August 2001), <<http://www.arl.org/newsltr/217/mhp.html>>.

Id., *Check out the new library: an interview with Clifford Lynch*, "Ubiquity", 4 (2003), 23, <http://www.acm.org/ubiquity/interviews/pf/c_lynch_1.html>.

Id., *Institutional repositories: essential infrastructure for scholarship in the digital age*, "ARL Bimonthly Report", (2003), 226, <<http://www.arl.org/newsltr/226/ir.html>>.

MARCO MARANDOLA, *Il diritto d'autore in ambito digitale secondo la legislazione italiana: un problema aperto*, "Storiadelmondo", (2004), 2, <<http://www.storiadelmondo.com/22/marandola.digitalizzazione.pdf>>.

Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), <<http://www.openarchives.org/OAI/openarchivesprotocol.htm>>.

VANESSA SPEDDING, *Great data, but will it last?*, "Research Information", (2003), 5, <<http://www.researchinformation.info/rispring03data.html>>.

ROBERT TANSLEY et al., *DSPACE: an open source dynamic digital repository*, "D-Lib Magazine", 9 (2003), 1, <<http://www.dlib.org/dlib/january03/smith/01smith.html>>.

ALICE TARLETON, *Cambridge first with electronic archive*, "Guardian Unlimited", 11 February 2003, <<http://education.guardian.co.uk/higher/news/story/0,9830,893535,00.html>>.

The DSPACE Project, "MIT Faculty Newsletter", 12 (2000), 4, <<http://DSpace.org/news/articles/DSpace-project.html>>.

HERBERT VAN DE SOMPEL – CARL LAGOZE, *Notes from the interoperability front: a progress report on the Open Archives Initiative*, 6th European Conference on Research and Advanced Technology for Digital Libraries, September 2002, Rome, <<http://www.openarchives.org/documents/ecdl2002-oai.pdf>>.

GIUSEPPE VITIELLO, *L'identificazione degli identificatori*, "Biblioteche oggi", 22 (2004), 2, p. 67-80, <<http://www.bibliotecheoggi.it/2004/20040206701.pdf>>.

PAUL WHEATLEY, *A way forward for developments in the digital preservation functions of DSPACE: options, issues and recommendations*, July 25, 2003, <<http://DSpace.org/news/articles/DpAndDSpace.pdf>>.

ANN WOLPERT, *The future of electronic data*, "Nature", (2002), 420, p. 17, <<http://DSpace.org/news/articles/future.pdf>>.

JEFFREY R. YOUNG, *"Superarchives" could hold all scholarly output*, "The Chronicle of Higher Education", July 5, 2002, <<http://chronicle.com/free/v48/i43/43a02901.htm>>.

Note

- ¹ CLIFFORD A. LYNCH, *Institutional repositories: essential infrastructure for scholarship in the digital age*, "ARL", (February 2003), 226, p. 1-7, <<http://www.arl.org/newsltr/226/ir.html>>.
- ² Come sta facendo ad esempio l'iniziativa OpenCourseWare presso il MIT, <<http://ocw.mit.edu>> (ultima visita: 3 aprile 2004).
- ³ <<http://www.openarchives.org/meetings/SantaFe1999/ups-invitation-ori.htm>>.
- ⁴ <<http://www.openarchives.org>>.
- ⁵ <<http://www.soros.org/openaccess/>>.
- ⁶ HERBERT VAN DE SOMPEL – CARL LAGOZE, *The Santa Fe Convention of the Open Archives Initiative*, "D-Lib Magazine", (February 2000), 6.
- ⁷ SALLY ATWOOD, *MIT's superarchive*, "Technology Review", December 2002/January 2003, <<http://www.technologyreview.com/articles/atwood1202.asp>> (ultima visita: 3 marzo 2004).
- ⁸ *Ibidem*.
- ⁹ ROBERT TANSLEY et al., *DSPACE: an open source dynamic digital repository*, "D-Lib Magazine", 9 (2003), 1, <<http://www.dlib.org/dlib/january03/smith/01smith.html>> (ultima visita: 18 febbraio 2004).
- ¹⁰ *Ibidem*.
- ¹¹ HOWARD BESSER, *The next stage: moving from isolated digital collections to interoperable digital archives*, "First Monday", 7 (2002), 6, <http://firstmonday.org/issues/issue7_6/besser/index.html> (ultima visita: 22 marzo 2004).
- ¹² <<http://www.dspace.org/federation/>>.
- ¹³ ALICE TARLETON, *Cambridge first with electronic archive*, "Guardian Unlimited", 11 February 2003, <<http://education.guardian.co.uk/higher/news/story/0,9830,893535,00.html>> (ultima visita: 6 marzo 2004).
- ¹⁴ ERIC CELESTE – MARGRET BRANSCHOF-SKY, *Building DSPACE to enhance scholarly communication*, in *E-Serials: publishers, libraries, users and standards*, 2nd ed. Wayne Jones (ed.),

Binghamton, NY, The Haworth Press, 2003, p. 239-247, <<http://libraries.mit.edu/dspace-mit/news/articles/celeste.pdf>> (ultima visita: 17 marzo 2004).

¹⁵ STEVE LAWRENCE, *Online or invisible?*, "Nature", 411 (2001) 6837, p. 521. Versione online: <<http://www.neci.nec.com/~lawrence/papers/online-nature01/>> (ultima visita: 17 marzo 2004).

¹⁶ Una valida discussione sulla trasformazione della peer-review nel nuovo panorama di disseminazione ad accesso libero è disponibile alla URL: <<http://www.ecs.soton.ac.uk/~harnad/Hypermail/Amsci/0479.html>>.

¹⁷ *Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)*, <<http://www.openarchives.org/OAI/openarchivesprotocol.htm>> (ultima visita: 4 marzo 2003).

¹⁸ <<http://dspace.org/technology/metadata.html>>.

¹⁹ L'articolo recita: "Sono protette ai sensi di questa legge le opere dell'ingegno di carattere creativo che appartengono alla letteratura, alla musica, alle arti figurative, all'architettura, al teatro ed alla cinematografia, qualun-

que ne sia il modo o la forma di espressione".

²⁰ Art. 118: "Il contratto con il quale l'autore concede a un editore l'esercizio del diritto di pubblicare per le stampe, per conto e a spese dell'editore stesso, l'opera dell'ingegno, è regolato, oltreché dalle disposizioni contenute nei codici, dalle disposizioni generali di questo capo e dalle disposizioni particolari che seguono".

²¹ Art. 71 ter: "È libera la comunicazione o la messa a disposizione destinata a singoli individui, a scopo di ricerca o di attività privata di studio, su terminali aventi tale unica funzione situati nei locali delle biblioteche accessibili al pubblico, degli istituti di istruzione, nei musei e negli archivi, limitatamente alle opere o ad altri materiali contenuti nelle loro collezioni e non soggetti a vincoli derivanti da atti di cessione o da licenza".

²² <<http://www.creativecommons.org>>.

²³ FELICIANA FAIELLA, *Digital preservation*, "Biblioteche oggi", 18 (2000), 8, p. 64-69.

²⁴ In un deposito digitale l'interscambio di informazioni, in entrata e in

uscita da esso, avviene attraverso l'utilizzazione di pacchetti di informazione (IP, Information Packages), che fungono da contenitori di dati. Gli IP sono di tre tipi, a seconda che le informazioni provengano dal produttore dell'OAI, all'interno del deposito o dal deposito all'utente: SIP, Submission Information Package (Pacchetto di informazioni per l'immissione), utilizzato per l'inserimento dei dati nell'archivio; AIP, Archival Information Package (Pacchetto di informazioni per l'archiviazione), finalizzato alla conservazione sul lungo periodo; DIP, Dissemination Information Package (Pacchetto di informazioni per la disseminazione), accessibile direttamente dall'utente finale. Per la disseminazione (DIP) DSpace utilizza un particolare schema XML, nel quale vengono esportati i metadati, ma è in corso una collaborazione con la comunità del METS (Metadata Encoding & Transmission Standard), al fine di sviluppare uno schema più completo ed esteso.

²⁵ Ulteriori dettagli sul DOI sono disponibili all'indirizzo: <<http://www.doi.org>>.