

Le linee guida per l'accesso aperto alle tesi di dottorato

STEFANIA ARABITO, DANIELA CERMESONI,
PAOLA GALIMBERTI, MARIALAURA VIGNOCCHI

Pur essendo prodotti della ricerca a tutti gli effetti, le tesi di dottorato hanno sempre avuto una scarsa diffusione e un sottoutilizzo derivato dal fatto che risultano consultabili solo presso le Biblioteche Nazionali Centrali di Roma e Firenze, dove l'accesso e la fruibilità sono limitati da una serie di vincoli. Il trattamento delle tesi, inoltre, è risultato essere molto difforme fra le diverse realtà universitarie. Accanto a realtà più avanzate dove le tesi sono già disponibili ad accesso aperto attraverso un deposito istituzionale, ci sono realtà nelle quali le tesi vengono raccolte in formato cartaceo. Il gruppo di lavoro CRUI OA si è posto come primo obiettivo la stesura di Linee guida che servissero come punto di riferimento agli Atenei italiani, perché potessero essere adottate soluzioni simili e coerenti. L'articolo illustra i punti fondamentali delle Linee guida per l'accesso aperto alle tesi di dottorato, soffermandosi soprattutto sulle problematiche legate al quadro normativo, ai metadati e ai formati.

Parole chiave: CRUI – Depositi istituzionali – ETD – Open Access – Tesi di dottorato.

Il Gruppo di lavoro CRUI OA e le Linee guida

La nascita di un Gruppo di lavoro che si occupasse dell'*Open Access* [1] è stata voluta nell'aprile 2006 dalla Commissione Biblioteche della Conferenza dei Rettori delle Università Italiane (CRUI), presieduta dal Prof. Vincenzo Milanese, che ne affida il coordinamento al Prof. Roberto Delle Donne [2].

Scopo del Gruppo, che si è avvalso del supporto dell'avv. Matelda Grassi, è di attuare i principi della *Dichiarazione di Berlino per l'accesso aperto alla letteratura scientifica* [3], sostenuta dalla *Dichiarazione di Messina* [4], con le quali si promuove Internet quale strumento funzionale alla conoscenza scientifica generale di base e alla speculazione umana, e si indicano le misure che le figure dominanti nelle politiche di ricerca, le istituzioni scientifiche, i finanziatori, le biblioteche, gli archivi ed i musei devono tenere in considerazione.

In particolare, il sottogruppo che si occupa delle tesi di dottorato si propone come obiettivo la stesura di Linee guida per il deposito delle tesi di dottorato negli archivi

aperti [5], che vengono approvate dalla CRUI nel novembre 2007, e immediatamente inviate ai Rettori di tutti gli Atenei italiani. Le Linee guida nascono dalla necessità di individuare uno *standard* nazionale nella raccolta ed esposizione dei dati e presentano una serie di raccomandazioni rivolte agli Atenei affinché possano realizzare una procedura di deposito delle tesi di dottorato negli archivi istituzionali autonoma, ma nel contempo basata su soluzioni simili e coerenti. Per facilitare il lavoro tecnico e amministrativo, sono allegati alle Linee guida una proposta di modifica del regolamento dei bandi di dottorato, la declaratoria da far firmare ai dottorandi già in corso, lo schema dei metadati da adottare per gli archivi aperti, una serie di indicazioni pratiche per i dottorandi relative al diritto d'autore.

Perché le tesi di dottorato

Le tesi di dottorato hanno una doppia natura. Da un lato sono documenti amministrativi che certificano la regolarità del dottorato conseguito e che pertanto vanno conservate nel fascicolo dello studente per un periodo di tempo illimitato, dall'altro sono documenti bibliografici, che attestano l'attività scientifica svolta dal dottore di ricerca e quindi, secondo la legge, devono essere pubblicamente accessibili, e secondo i principi della Berlin Declaration è opportuno che lo siano.

Questo secondo aspetto è molto interessante e consente di riflettere sul valore scientifico delle tesi [6]. Esse offrono una dissertazione dell'argomento trattato più ampia che negli articoli scientifici, garantendo un'analisi bibliografica retrospettiva più approfondita. Il loro *status* di validazione può essere paragonato a quello dei *preprint*, in quanto non subiscono un processo formale di *peer-review*, sebbene l'opera di supervisione e tutorato del docente di riferimento sia molto approfondita e prolungata nel tempo. Inoltre, l'autorità che convalida la tesi (la commissione giudicatrice, il collegio dei docenti, il docente di riferimento in primo luogo) è indicata pubblicamente e non secretata come avviene nei casi di *peer-review*: è quindi anch'essa sottoposta a un giudizio implicito nel momento in cui la tesi viene pubblicata sul web; pertanto, l'accuratezza nella certificazione della validità scientifica della tesi non può venire meno [7].

Le tesi di dottorato in Italia non si concludono con una pubblicazione come in altre parti d'Europa, esse non sono formalmente pubblicate, non devono essere richiesti permessi agli editori, gli autori non perdono alcun profitto acconsentendo all'*Open Access*.

Le Università sono nella posizione di poter prevedere il deposito obbligatorio delle tesi (come sottolinea Peter Suber, «mandates work and exhortations don't» [8]).

Le tesi di dottorato hanno attualmente scarsa visibilità, nascoste nel web profondo, talora rintracciabili solo attraverso cataloghi cartacei. Se si pensa che le tesi nascono già digitali, e che la loro discussione è pubblica, il deposito in archivi istituzionali ad accesso

aperto non può essere se non un prolungamento della loro vita, aumentando l'impatto sia a vantaggio del dottore di ricerca che dell'istituzione cui appartiene.

La situazione italiana in materia di accesso aperto: il ritardo rispetto all'Europa

Al fine di raccogliere informazioni sulla situazione delle tesi in formato elettronico in Italia, nel mese di novembre 2005 è stata effettuata a cura di vari esponenti europei una prima indagine, aggiornata poi alla fine del 2006 [9]. L'analisi si focalizza sulle modalità di raccolta e sul *workflow*, e, sebbene abbiano risposto solo 19 Atenei, evidenzia che la situazione italiana è frammentaria, difforme, priva di *standard* e buone pratiche di riferimento. Pochi Atenei [10], hanno già una prassi consolidata di raccolta delle tesi in formato elettronico, altri sentono l'esigenza, per motivi di razionalizzazione e di visibilità, di raccogliere questi documenti nati in formato digitale, ma mancano delle *policy* nazionali. Ogni Università adotta politiche proprie rispetto al meccanismo di consegna per il deposito alle Biblioteche Nazionali, rispetto alla maggiore o minore interazione con le segreterie didattiche o con quelle di dottorato, rispetto ai formati di consegna e ai metadati, mentre restano del tutto in sospeso le questioni legate al diritto d'autore.

La produzione italiana conta un numero crescente di tesi: nel 2000 sono state discusse circa 4.000 tesi, diventate 9.800 nel 2005. La produzione media annua per Ateneo va dalle 50 tesi negli Atenei più piccoli alle 500/800 in quelli più grandi, con una media fra le 200 e le 300 tesi. Si tratta di un patrimonio notevole, rimasto fino ad ora pressoché invisibile.

D'altra parte, nella situazione internazionale la raccolta di tesi in archivi istituzionali e/o nazionali è una pratica diffusa e che vediamo rappresentata in diversi progetti, di cui si espongono i più significativi.

Promise of Science [11]

Le tesi di dottorato delle Università olandesi (la maggior parte delle quali prevede l'obbligo di deposito nell'archivio istituzionale) vengono catturate mediante meccanismi di *harvesting* e rese disponibili attraverso il sito del progetto Promise of Science. Il sito raccoglie (ad oggi) più di 19.000 tesi di dottorato, la maggior parte consultabili in full-text.

Ethos [12]

In Gran Bretagna il progetto EThOS (Electronic Theses Online Service) in cui sono confluite le esperienze dell'Università di Glasgow (Daedalus) e di Thesis Alive! [13]

(Cambridge, Cranfield, Leeds, Manchester) si propone di fornire un'infrastruttura e un supporto per *l'harvesting* delle tesi di dottorato delle Università consorziate anche in accordo con la British Library, in modo da dare la maggiore visibilità possibile a questo tipo di materiali. Il progetto ha avuto ingenti finanziamenti da JISC [14] e da CURL [15].

Dissonline [16]

La Deutsche Nationalbibliothek raccoglie da anni le tesi di dottorato in formato elettronico con un sistema di *harvesting* dagli archivi istituzionali delle Università tedesche. Le tesi possono essere consultate in Internet direttamente dal portale della Biblioteca Nazionale. Attualmente i documenti in *full-text* disponibili sono più di 45.000.

Australasian Digital Theses [17]

In Australia con ADT è stato creato un *database* nazionale distribuito il cui scopo è di fornire uno *standard* per il deposito e l'archiviazione delle tesi. Ogni Università che partecipa al progetto archivia le proprie tesi in un *server* locale usando però la medesima configurazione del *database* centrale e gli stessi metadati in modo che sia garantita l'interoperabilità. Il formato obbligatorio per i documenti è il PDF. Le tesi possono essere ricercate sia attraverso l'interfaccia ADT sia localmente. Il *database* contiene attualmente 23.600 tesi in formato digitale.

European E-theses [18]

Il progetto, finanziato da JISC (UK), Surf (NL) e DIVA (SW), è iniziato nel 2006 e terminato con l'ETD 2007 [19]. Scopo del Gruppo di coordinamento del progetto (*GUIDE working group*) era di fornire linee guida e raccomandazioni a livello tecnico e legale per l'archiviazione, l'accesso e *l'harvesting* delle tesi di dottorato a livello europeo. Hanno partecipato al progetto Gran Bretagna, Olanda, Germania, Danimarca e Svezia.

Dart Europe [20]

È un portale di accesso alle tesi di dottorato di Università e consorzi di Università europee. Permette l'accesso ad oltre 81.400 tesi di dottorato di 33 *partner*. Dart è il *partner* europeo di NDLTD. Più volte è stata richiesta la partecipazione dell'Italia al progetto.

Il Gruppo CRUI OA, poiché il ritardo dell'Italia rispetto all'Europa in questo ambito è molto forte, ha preferito concentrarsi sull'accessibilità delle tesi in quanto documenti bibliografici e prodotti della ricerca scientifica, senza però perdere di vista la natura amministrativa delle tesi. A questo proposito, il Gruppo di lavoro è già in contatto

con gli archivisti al fine di elaborare una strategia comune che, accanto all'accessibilità, possa garantire la tutela dell'autenticità e dell'integrità di questi documenti.

Il complesso quadro normativo

La raccolta, conservazione ed esposizione delle tesi di dottorato (in qualsiasi formato), è regolata da una serie di norme di origine diversa delle quali è necessario tenere conto, che sono di interpretazione non univoca e spesso in contraddizione fra loro.

La legge 22 aprile 1941, n. 633 (legge sul diritto d'autore) e successive modificazioni, non menziona le tesi tra i documenti tutelati. Non esiste praticamente giurisprudenza in merito a chi debba essere considerato l'autore della tesi, tranne la citatissima sentenza della Corte d'Appello di Perugia del 22 febbraio 1995 che attribuisce allo studente il diritto di paternità e quello di sfruttamento economico dell'opera [21]. È importante ricordare che la legge sul diritto d'autore tutela la forma, non l'idea (che potrebbe anche essere stata suggerita dal docente *tutor*). Ciò che viene protetto non è dunque l'idea rappresentata, ma la particolarità e creatività della forma espositiva, che dovrebbero essere frutto del lavoro autonomo del dottorando.

Il documento fondamentale che regola la consegna e conservazione delle tesi di dottorato è quello che istituisce il dottorato di ricerca (Legge 21 febbraio 1980, n. 28 "Legge delega per l'istituzione del dottorato di ricerca"), in cui si dice:

Il rilascio del titolo di dottore di ricerca è subordinato al deposito di copie, anche non stampate, dei lavori sulla base dei quali il titolo è stato conseguito presso le Biblioteche nazionali di Roma e Firenze, che ne devono assicurare la pubblica consultabilità per non meno di trenta anni. I testi di cui sopra devono essere corredati dalla relazione dei commissari, incluse le eventuali relazioni di minoranza.

La legge viene poi modificata dal D.M. n. 224 del 30/04/99: regolamento in materia di dottorato di ricerca, in particolare nel seguente punto:

Il titolo è rilasciato dal rettore dell'Università che, a richiesta dell'interessato, ne certifica il conseguimento. Successivamente al rilascio del titolo, l'Università medesima cura il deposito di copia della tesi finale presso le biblioteche nazionali di Roma e Firenze.

Nel decreto ministeriale non si parla più di conservazione per 30 anni. E non si parla nemmeno più di pubblica consultabilità. Poiché però *mission* delle Biblioteche Nazionali è la raccolta e conservazione dei prodotti culturali nazionali, si è intesa l'omissione della limitazione temporale come indicazione per una conservazione a lungo termine, senza scadenza alcuna.

La Legge 15 aprile 2004, 106 "Norme relative al deposito legale dei documenti di

interesse culturale destinati all'uso pubblico" (GU 98 del 27 aprile 2004), che introduce la riforma del deposito legale con l'istituzione di archivi di conservazione regionali e nazionali, prevede l'estensione dell'obbligo del deposito legale ai documenti in formato elettronico e diffusi tramite rete telematica quali ad esempio le tesi di dottorato se conservate negli archivi istituzionali. Il regolamento attuativo della nuova legge, DPR 3 maggio 2006, n. 252 (GU 191 del 18 agosto 2006), pur rinviando le modalità di deposito dei documenti diffusi via Internet a un ulteriore decreto attuativo, promuove forme volontarie di sperimentazione del deposito mediante la stipula di appositi accordi con i soggetti obbligati al deposito *che definiranno anche le modalità tecniche del deposito, prevedendo ove possibile, anche forme automatiche di raccolta, secondo le migliori pratiche e conoscenze internazionali del settore.*

A complicare il quadro è intervenuta la circolare ministeriale n. 1420 del 28 luglio 2006, che prevede la consegna delle tesi di dottorato in formato elettronico tramite invio alle BNC [Biblioteche Nazionali Centrali] dei *file* su CD o DVD. Il gruppo di lavoro OA CRUI, in accordo con le BNC, preoccupate delle possibili conseguenze della circolare ministeriale del 2006 per la conservazione a lungo termine delle tesi, ha cercato di promuovere la collaborazione fra Università e BNC al fine di avviare una forma sperimentale di deposito legale tramite *harvesting* dei file via web. Purtroppo la circolare ministeriale n. 1746 del 20 luglio 2007, che invita le Università alla consegna delle tesi in formato elettronico tramite *harvesting* o caricamento sicuro dei file direttamente sul sito delle Biblioteche Nazionali, rettifica le indicazioni precedenti alludendo troppo sinteticamente all'accordo con le BNC, responsabili del deposito legale e della conservazione a lungo termine. Per la loro sinteticità e la loro contraddittorietà, le circolari hanno preoccupato la componente archivistica degli atenei italiani impegnati nello studio dei formati e dei metadati più adatti alla conservazione anche a lungo termine del documento amministrativo.

Diritti e tesi di dottorato

Se si parte dall'assunto che autore della tesi è il dottorando, a lui spettano tutti i diritti morali e di sfruttamento economico.

La messa *online* delle tesi di dottorato implica una serie di azioni da parte degli Atenei che allo stato attuale vanno autorizzate dall'avente diritto [22]. Le Linee guida prevedono dunque che gli Atenei modificchino i bandi di dottorato e i regolamenti provvedendo a disciplinare l'inserimento nei propri archivi aperti e il deposito legale presso le Biblioteche Nazionali. L'accettazione da parte del dottorando della borsa di studio implicherà anche l'accettazione di tutte le condizioni poste nel bando. Chi non fosse d'accordo su uno o più punti, per esempio sulla pubblica accessibilità del prodotto

finale, dovrà rinunciare al dottorato.

Per il periodo che intercorre da qui alla entrata in vigore dei nuovi regolamenti, gli Atenei dovranno individuare soluzioni transitorie; consiglieranno ai dottorandi di archiviare la tesi nel deposito istituzionale, presentando loro tutti i vantaggi in termini di impatto e visibilità della loro ricerca, ma non potranno imporre nessun obbligo, in quanto senza la modifica del bando è l'autore l'unico a poter decidere cosa fare della propria opera. Resta il fatto che la tesi sarà comunque pubblicamente consultabile negli *edifici* delle Biblioteche Nazionali.

Le Linee guida prevedono pochi, motivati casi nei quali la tesi può essere secretata per un periodo di tempo determinato (*embargo*) [23], pur restando fermo l'obbligo di deposito immediato dopo la discussione della tesi:

- se la tesi è in corso di pubblicazione e il dottorando ha già un accordo in corso con un editore che non accetta una pregressa pubblicazione ad accesso aperto, è possibile posticipare l'accesso [24];
- se la tesi è frutto di un progetto di ricerca finanziato da un ente esterno con il quale esiste un contratto pregresso che impone il vincolo di segretezza, può essere richiesto un periodo di embargo per proteggere il valore commerciale delle nuove informazioni;
- se la tesi ha per argomento un personaggio in vita o da poco deceduto per il quale si teme di violare il diritto alla *privacy*, la tesi può essere oggetto di secretazione.

Brevetti

Un discorso a parte meritano i brevetti. Molti Atenei considerano la discussione della tesi di dottorato una forma di pre-divulgazione. Ciò va contro al requisito fondamentale per la richiesta di un brevetto che è la novità del trovato che non deve essere mai stato comunicato in pubblico in alcuna forma. La richiesta di brevetto deve quindi essere necessariamente inoltrata prima della discussione della tesi in quanto la discussione equivale ad una pubblicazione. Non è l'accesso aperto alla tesi a pregiudicare la richiesta di brevetto, ma la discussione stessa.

Dopo aver depositato la domanda e discusso la tesi, sarà possibile depositarla nell'archivio ad accesso aperto.

Utilizzo di materiali di terzi

In ambito analogico capitava spesso che il dottorando inserisse nella propria tesi foto, grafici, immagini, parti di testo (anche di propri articoli) sui quali gravavano diritti

di terzi. Tale prassi, tollerata anche se scorretta, non è riproducibile in ambito digitale, perché il tipo di circolazione e di diffusione cambia.

Il Gruppo di lavoro ha prodotto un allegato alle Linee guida [25] in cui per le varie tipologie di materiale si spiega ai dottorandi come comportarsi nel caso di materiali di terzi.

In generale, possiamo dire che è sempre possibile l'utilizzo di materiali di pubblico dominio o per i quali si è chiesto e ottenuto un permesso scritto da parte dell'avente diritto. Senza permesso non è possibile utilizzare materiali di terzi, a parte le foto e le immagini, che in forza dell'articolo 70.1bis possono essere liberamente utilizzate in internet purché degradate [26].

Misure tecnologiche di protezione

È un luogo comune l'idea che la pubblicazione in Internet faciliti il plagio: mentre, in effetti, la pubblicazione in Internet rende più facile individuare il plagio. Inoltre il maggior deterrente in ambito accademico è la riprovazione della comunità scientifica di riferimento.

Le tesi di dottorato devono essere auto-archivate nei depositi istituzionali, prive di misure tecnologiche di protezione o di sistemi di DRM. Se viene richiesto un embargo sarà il responsabile dell'archivio a rendere la tesi non consultabile. La presenza di misure tecnologiche di protezione, aggirabili solo con difficoltà, rende infatti il documento non idoneo ai fini della conservazione a lungo termine.

L'*harvesting* delle tesi di dottorato a cura delle BNC

A seguito della circolare ministeriale n. 1746 del 20 luglio 2007 e in funzione del regolamento attuativo della nuova legge sul deposito legale, le BNC di Firenze e di Roma si sono rese disponibili a realizzare una procedura di deposito legale delle tesi di dottorato in formato elettronico attraverso la raccolta automatizzata (*harvesting*) dei metadati e del *full-text*. L'iniziativa rappresenta per le BNC un'occasione per implementare e testare gli obiettivi programmatici del più ampio progetto "Magazzini Digitali", il cui scopo è la realizzazione di depositi digitali accreditati e certificati per la conservazione a lungo termine delle risorse diffuse attraverso reti telematiche [27]. La nuova procedura appare migliorativa rispetto a quella proposta nella precedente circolare ministeriale che autorizzava la consegna delle tesi su CD o DVD, esponendo gli istituti deputati alla conservazione a lungo termine alle innumerevoli problematiche legate alla gestione di *file* su supporto fisico senza tuttavia creare le premesse per un

reale miglioramento dei servizi bibliografici. Il deposito legale delle tesi attraverso *harvesting* automatizzato consente una semplificazione delle pratiche amministrative a carico degli atenei e una riduzione dei tempi legati alla catalogazione e alla gestione bibliografica a carico della Bibliografia Nazionale Italiana, rendendo possibile la realizzazione di un servizio nazionale di ricerca e accessibilità delle tesi di dottorato aggiornato ed efficiente.

La procedura di raccolta automatizzata si rivolge principalmente a tesi digitali contenute in *institutional repository* di ateneo compatibili con i più diffusi *standard* di interoperabilità. Per gli atenei che non hanno ancora implementato un *repository* compatibile con il protocollo OAI [28] è prevista la realizzazione di un modulo web che consente l'*upload* sicuro dei *file* e dei metadati delle tesi. Una volta archiviate le tesi di dottorato in un *institutional repository*, le successive fasi delle quali si compone la procedura di deposito legale sono tutte a carico delle BNC. Ecco in sintesi le 6 fasi della procedura [29]:

- 1) raccolta dei metadati secondo il OAI-PMH;
- 2) raccolta del *full-text* a partire dallo URL [Uniform Resource Locator] della tesi mediante un *crawler* che registra anche i metadati della cattura;
- 3) verifica dei formati, in particolare della loro traducibilità nel tempo (*renderability*). In particolare, un validatore automatico accerta che il formato dei *file* delle tesi sia un PDF/A, che in quanto standard ISO fornisce maggiori garanzie in termini di conservazione a lungo termine, così come raccomandato anche dal recente studio della Digital Preservation Coalition [30]. Se i file raccolti non sono PDF/A, i file vengono accettati con riserva e ne viene data informazione all'archivio di origine;
- 4) archiviazione delle tesi di dottorato (*full-text* e metadati) nel *trusted repository* delle BNC;
- 5) invio delle ricevute di avvenuto deposito tramite *e-mail* al gestore del *repository* di ateneo che contiene l'impronta digitale in formato SHA-1 delle tesi a garanzia dell'autenticità dei *file* archiviati;
- 6) accessibilità in rete delle tesi, che per il momento avviene solo presso le postazioni delle BNC, almeno fino a che non si sarà raggiunto un accordo sui metadati relativi alle licenze d'uso che possano essere trattati in modo automatizzato.

È stata completata una prima fase di *test* della procedura che ha interessato le tesi di dottorato contenute nell'archivio istituzionale AMS Tesi di Dottorato dell'Alma Mater Studiorum, Università di Bologna [31] e sono disponibili i primi dati sperimentali. L'*institutional repository* dell'Ateneo bolognese è un archivio dedicato in modo esclusivo al trattamento delle tesi di dottorato, separato dall'archivio destinato agli altri contributi di ricerca, a causa dei consistenti interventi di personalizzazione che sono stati effet-

tuati sul *software* EPrints 3.X per realizzare un'integrazione funzionale dell'archivio con il sistema gestionale dei dottorati dell'Ateneo. Avviato verso la fine del 2006, contiene circa 561 tesi che corrispondono alla produzione del 2007. Sono stati implementati per il momento 2 *set* di metadati mappati l'uno sull'altro: uno più ricco per la gestione e la fruizione locale, l'altro limitato agli elementi del Dublin Core semplice esposto secondo l'OAI-PMH per il deposito legale con le BNC.

La sperimentazione della procedura di deposito, conclusasi con successo con l'invio delle impronte digitali a conferma dell'avvenuto deposito, ha fatto emergere alcune prevedibili criticità soprattutto per quanto concerne i formati dei *file*. Dal momento che non si era riusciti a mettere a disposizione dei dottorandi un adeguato supporto tecnico e anche per introdurre gradualmente i cambiamenti e le modifiche di procedura, si è scelto di non imporre la consegna dei *file* in PDF/A, limitandosi a richiedere genericamente la tesi in un *file* PDF. La validazione a cura delle BNC ha comunque rilevato che quasi il 18 % dei *file* raccolti sono in un formato compatibile con lo *standard* ISO PDF/A. Per il prossimo anno sono già state pianificate una campagna di sensibilizzazione e l'implementazione di strumenti in rete per la realizzazione di formati compatibili. La procedura di auto-archiviazione dell'IR di Ateneo verrà poi corredata di un validatore che segnalerà al dottore di ricerca eventuali difformità dei formati dei *file* caricati o la presenza di protezioni.

La sperimentazione della procedura da parte delle BNC è ancora in corso. I prossimi interventi riguarderanno prove di *harvesting* da archivi "misti" in cui le tesi di dottorato coesistono insieme con gli altri prodotti della ricerca dell'ateneo e una migliore definizione e gestione dei metadati relativi ai diritti.

Condivisione dei metadati

Le raccomandazioni del Gruppo di lavoro CRUI OA sull'accesso aperto si basano su un'analisi comparativa dei *set* di metadati utilizzati negli archivi istituzionali in ambito italiano ed europeo. In allegato alle Linee guida si trova infatti uno schema condiviso che costituisce il punto di partenza per l'interoperabilità degli archivi e per la standardizzazione delle procedure, riducendo al minimo la necessità di successivi interventi per la normalizzazione dei metadati.

Allo stato attuale, l'unico obiettivo realistico è arrivare al primo livello di interoperabilità, condividendo un protocollo comune per lo scambio dei dati che renda possibile l'*harvesting* dei *full text* delle tesi di dottorato con una descrizione accurata dei campi e delle etichette. In prospettiva, si ambisce a costruire servizi a valore aggiunto sugli archivi.

Il Dublin Core semplice non è sufficiente per eseguire ricerche avanzate, che sono

invece rese possibili dall'adozione del Dublin Core qualificato.

Di seguito i metadati considerati obbligatori:

- 1) *dc.title*: titolo;
- 2) *dc.creator*: autore (cognome, nome);
- 3) *dc.description: abstract* (meglio se in inglese);
- 4) *dc.language*: lingua (formato ISO639-1);
- 5) *dc.identifier*: URL del *full-text* della tesi o pagina di passaggio;
- 6) *dc.type: Doctoral Thesis* (solo in inglese);
- 7) *dc.contributor: tutor*/relatore (cognome, nome);
- 8) *dc.date*: data di pubblicazione (ISO 8601), i.e. date della discussione (evitare altre date che potrebbero indurre confusione);
- 9) *dc.publisher*: nome dell'Università;
- 10) *dc.format*: dimensioni in bytes/MIME type.

Tra i metadati fortemente raccomandati:

- 11) *dc.subject*: settori scientifico-disciplinari MIUR;
- 12) *dc.rights*: embargo o consultabilità immediata.

Gli archivi che adotteranno questo schema di metadati saranno compatibili con gli *standard* europei. Sarebbe auspicabile arrivare alla condivisione del contesto informativo più ampio, in linea con il Processo di Bologna e più in particolare con la cosiddetta Convenzione di Lisbona (Convenzione sul riconoscimento dei titoli di studio relativi all'insegnamento superiore nella regione europea) [32].

Formati

L'aspetto dei formati rimane un punto controverso, specialmente per quanto riguarda le tesi in formato digitale. I requisiti da soddisfare sono svariati, considerando sia il valore bibliografico sia quello archivistico delle tesi di dottorato.

In quanto documenti amministrativi, bisogna assicurare e garantire la loro autenticità, integrità e fissità ai fini di una preservazione a lungo termine. Devono essere anche corredate dai metadati archivistici.

In quanto documenti bibliografici, devono soddisfare i criteri di accessibilità in rete.

I giovani ricercatori che le auto-archiviano non hanno però competenze specifiche in materia di conservazione digitale e in ambito informatico; per lo più sono in grado di redigere testi ed eseguire un'archiviazione di base. I dottorandi non hanno consapevolezza delle problematiche correlate all'obsolescenza dei formati, ma hanno interesse a proteggere il loro lavoro da modifiche (intenzionali o involontarie) dopo la pubblicazione in rete. Tuttavia, come abbiamo visto, i *file* non devono essere criptati altrimenti non è possibile effettuare la *refresh* all'occorrenza.

In realtà, sarebbe opportuno formare laureandi e dottorandi sui temi specifici della preservazione a lungo termine e fornire loro gli strumenti per produrre file XML corredati da metadati bibliografici e archivistici: questo risolverebbe gli aspetti archivistici e biblioteconomici del problema. Sarebbe inoltre conveniente collaborare in modo più efficace con i docenti e fornire loro degli strumenti di *authoring* efficaci ed amichevoli, idealmente in modo trasparente.

È necessaria una strategia a livello nazionale. Tutte le componenti coinvolte, gli archivisti e le Biblioteche Nazionali, dovrebbero contribuire con le proprie competenze specifiche nel campo della preservazione a lungo termine. Di fatto, gli archivisti italiani stanno elaborando un *set* di metadati specifici, d'intesa con le Biblioteche Nazionali, che hanno per statuto il compito di preservare e garantire l'accesso al patrimonio bibliografico-documentale italiano.

È impossibile in questa fase dire l'ultima parola sui formati. Eppure le Università italiane devono cominciare a raccogliere e conservare le tesi in formato digitale.

La scelta del PDF/A – conformemente alle richieste delle Biblioteche Nazionali – non è stata facile e condivisa da tutti; il PDF è considerato il male assoluto da molti – chiunque fosse presente ad Open Repositories 2008 a Southampton ha ben presente la questione – eppure è diventato uno *standard de facto* in tutta Europa e oltre (fa eccezione la Germania). Le recenti raccomandazioni della Digital Preservation Coalition rafforzano questa posizione, anche se la questione rimane assolutamente aperta e bisogna tenere alta l'attenzione.

Tesi digitali e archivi istituzionali: lavori in corso

Lo scorso gennaio il Gruppo CRUI OA ha somministrato un questionario alle Università italiane. Questi i risultati dell'indagine.

- 25 Università raccoglievano o stavano per raccogliere le tesi di dottorato in formato digitale nei loro archivi istituzionali, per lo più per renderle disponibili ad accesso aperto;
- La produzione locale di tesi di dottorato varia da 50 a 500-800 l'anno secondo le dimensioni dell'Ateneo; la media è di 200-300 tesi l'anno;
- Sono pochi i Senati accademici che hanno ufficialmente deliberato il deposito delle tesi di dottorato negli archivi istituzionali; la metà degli Atenei ha una politica mandataria solo per il deposito, l'altra metà ha reso obbligatorio l'accesso in *Open Access*;
- I principali promotori del deposito negli archivi istituzionali sono i bibliotecari (con qualche eccezione);
- Le procedure relative all'archiviazione, validazione e pubblicazione sono a cura dei

bibliotecari (con qualche eccezione), con la collaborazione del personale amministrativo;

- DSpace ed Eprints sono i *software* più utilizzati. Gli archivi istituzionali sono nella maggior parte dei casi integrati con altre banche dati, soprattutto amministrative, con i sistemi di autenticazione, con gli archivi della ricerca, con gli OPAC, con sistemi di *cross-searching* e *link-resolving*;
- L'auto-archiviazione è la modalità di immissione dati consigliata e prevista dagli *iter* procedurali d'Ateneo; ciononostante, nella maggior parte dei casi sono i bibliotecari ad archiviare le tesi in formato digitale;
- Viene concessa la possibilità di un embargo variabile dai sei mesi ai tre anni. 12 mesi è il periodo *standard* di embargo per quasi la metà degli Atenei; raramente è il dottorando a decidere la durata dell'embargo; in ogni caso si richiede la sottoscrizione di una declaratoria in cui si fornisce la motivazione per l'embargo. I metadati sono disponibili e ricercabili da subito;
- Nella metà degli Atenei vengono forniti servizi aggiuntivi alla mera gestione delle tesi: *print on demand*, deposito legale, statistiche, consulenze legali, preservazione a lungo termine. L'altra metà degli Atenei sta per implementare servizi analoghi;
- Le biblioteche assicurano il supporto finanziario nella metà dei casi; nell'altra metà altri centri servizi dell'Ateneo contribuiscono alla manutenzione dell'archivio;
- Le attività ordinarie di gestione/manutenzione sono a cura dei bibliotecari, degli informatici e degli amministrativi in collaborazione con i docenti. È molto raro che i bibliotecari siano gli unici attori coinvolti;
- La metà degli archivi è in *outsourcing*, l'altra metà è gestita con risorse interne;
- L'interoperabilità a livello nazionale è considerata importante e necessaria da tutti. Si prendono in considerazione due livelli di interoperabilità: sintattica (ricerca federata su più archivi) e semantica (multilinguismo, per materia e per soggetto). Si auspica un servizio di *harvesting* dedicato a livello nazionale. Charamente, l'interoperabilità semantica è considerata un obiettivo molto più arduo da raggiungere.

Tutte le Istituzioni che hanno risposto al questionario considerano necessario affrontare una serie di questioni a livello europeo per valorizzare le tesi di dottorato:

- programmi unificati di archiviazione;
- adozione di *standard* comuni;
- attività costante di promozione e sensibilizzazione;
- progetti cooperativi nazionali;
- sistemi di formazione *post-lauream* equivalenti a livello italiano ed europeo;
- supporto a livello economico, legale e tecnico;

- valorizzazione del potenziale delle ricerche di dottorato;
- collaborazione con gli editori.

Conclusioni

L'obiettivo del Gruppo di lavoro CRUI OA era di pubblicare le Linee guida, come strumento di riferimento per le Università intenzionate ad archiviare le tesi di dottorato nei depositi istituzionali e renderle accessibili in *Open Access*. La condivisione degli stessi metadati è stato raccomandato con forza per ragioni di interoperabilità con i portali europei.

La finalità del lavoro era la disseminazione delle tesi di dottorato, considerata la loro importanza come prodotti della ricerca e la loro scarsa visibilità in rete con conseguente perdita di impatto. La certificazione e conservazione delle tesi di dottorato sono invece compiti cui altri organi sono preposti e di cui devono pertanto farsi carico, pur con la garanzia della massima collaborazione e disponibilità del Gruppo di lavoro.

Le Linee guida sono state ampiamente apprezzate da tutte le Istituzioni oggetto del sondaggio, come le risposte hanno dimostrato; le Università che avevano già istituito un deposito istituzionale le hanno considerate importanti nella misura in cui esse motivavano e appoggiavano la loro scelta; nella maggior parte dei casi, i vertici delle Istituzioni non hanno però dimostrato di essere pienamente consapevoli del potenziale del deposito, che non è stato ritenuto uno strumento istituzionale completamente legittimo.

Laddove invece il deposito delle tesi di dottorato era in fase di avvio, le Linee guida hanno fornito supporto al progetto e giustificato la richiesta di una *policy* istituzionale a sostegno dell'*Open Access*.

In ogni caso, le Linee guida rappresentano un passo fondamentale per portare a compimento a livello nazionale una strategia comune nei confronti dell'*Open Access*.

Naturalmente, la *mission* del Gruppo di lavoro non è stata ancora portata a termine. È stato intrapreso il primo passo, ma è necessaria una pressione incisiva per espandere i depositi istituzionali esistenti e per azzerare la resistenza al cambiamento. L'opposizione al deposito delle tesi di dottorato elettroniche negli archivi istituzionali e alla loro disponibilità in *Open Access* è ancora forte e diffusa.

Importante anche tenere contatti stretti e assidui con gli altri gruppi di lavoro europei ed internazionali per monitorare gli sviluppi e per essere aggiornati sulle loro attività.

I servizi a valore aggiunto per questo tipo di materiali necessitano anch'essi di una verifica a livello europeo: interoperabilità sintattica e semantica, l'uso di *standard* comuni, un *harvester* italiano dedicato e altri importanti temi sono i prossimi argomen-

ti d'analisi del Gruppo.

È di grande importanza anche poter informare i dottorandi sui loro diritti in qualità di autori e sul corretto utilizzo da parte loro di materiale di terzi.

Da ultimo, ma non per importanza, si auspica che lo scambio di esperienze con gli archivisti sulla conservazione a lungo termine e certificazione delle tesi di dottorato, al momento in fase iniziale, possa portare a una proficua e fattiva collaborazione.

Note

- [1] <www.cruui.it/HomePage.aspx?ref=894> consultato in data 2008-06-11. Esponenti del Gruppo sulle tesi di dottorato hanno poi partecipato con dei lavori a convegni internazionali: Paola Galimberti, Marialaura Vignocchi, *Time for a change: the Italian CRUI Open Access Working Group's action for a national e-theses provision service*, in "Proceedings ETD 2007 - Added value to e-theses. Uppsala (Sweden) 13th-16th June 2007". <eprints.rclis.org/archive/00012386> consultato in data 2008-06-11. Stefania Arabito, Daniela Cermesoni, Paola Galimberti, Marialaura Vignocchi, *Time to Harvest: Electronic Doctoral Theses in Italy*. in "Proceedings ETD2008: Spreading the Light: 11th International Symposium on Electronic Theses and Dissertations, Aberdeen (United Kingdom) 4yh-7th June 2008" <eprints.rclis.org/archive/00013743> consultato in data 2008-06-11. E convegni nazionali: Paola Galimberti, *Le Linee guida per l'accesso aperto alle tesi di dottorato*. in Serena Sangiorgi, Eds. *Proceedings Cristalli di esperienza: nuove prospettive e scenari per le tesi di dottorato: conservazione, accessibilità, certificazione, formati, integrazione con Open Access* "Quaderno CNBA" (10), pp. 9-20, Torino (Italy) <eprints.rclis.org/archive/00013742> consultato in data 2008-06-11
- [2] Roberto Delle Donne, *CRUI and Open Access in Italy*, in "Berlin 5 Open Access: From Practice to Impact: Consequences on Knowledge Dissemination Padova, Italia, 19-21 settembre 2007" <www.aepic.it/conf/viewpaper.php?id=311&cf=10> consultato in data 2008-06-11.
- [3] <oa.mpg.de/openaccess-berlin/berlin_declaration.pdf>; <oa.mpg.de/openaccess-berlin/singnatories.html> consultati in data 2008-06-11.
- [4] <www.aepic.it/conf/viewpaper.php?id=49&cf=1> consultato in data 2008-06-11.
- [5] <www.cruui.it/HomePage.aspx?ref=1149> consultato in data 2008-06-11.
- [6] Per un'ampia trattazione dei motivi che supportano il deposito delle tesi di dottorato negli archivi aperti istituzionali, si veda Peter Suber, *Open access to electronic theses and dissertations (ETDs)*, in "SPARC open access newsletter (SOAN)", July,2 2006 <www.earlham.edu/~peters/fos/newsletter/07-02-06.htm#etds> consultato in data 2008-06-11.
- [7] Il medesimo concetto è valido anche nel caso della pubblicazione da parte del personale docente della propria produzione scientifica (*pre/post print*) nei depositi istituzionali: solo il

materiale più significativo verrà esposto a pubblica lettura.

- [8] Cfr. Peter Suber, *supra*, nota 7.
- [9] La relazione relativa all'indagine 2006 si trova sul sito "European E-theses" al link <www.ukoln.ac.uk/repositories/european-e-theses/index/Italy> consultato in data 2008-08-05. Per i lavori preparatori si veda. www.surffoundation.nl/download/SMo_Survey.pdf consultato in data 2008-08-05. Susanna Mornati, *Le tesi di dottorato elettroniche in Europa*, "Bollettino del CILEA", 102 (2006), giugno, p. 53-53. bollettino.cilea.it/viewissue.php?id=53 consultato in data 2008-08-05.
- [10] Tra questi ricordiamo l'Università di Pisa, la Sapienza e il Politecnico di Milano (sebbene quest'ultimo non utilizzi il protocollo OAI-PMH).
- [11] <www.narcis.info/index/tab/proef/Language/EN> consultato in data 2008-06-11.
- [12] <www.ethos.ac.uk> consultato in data 2008-06-11.
- [13] <www.thesesalive.ac.uk> consultato in data 2008-06-11.
- [14] <www.jisc.ac.uk> consultato in data 2008-06-11.
- [15] <www.rluk.ac.uk> consultato in data 2008-06-11.
- [16] <www.dissonline.de/index.htm> consultato in data 2008-06-11.
- [17] <adt.caul.edu.au> consultato in data 2008-06-11.
- [18] <www.surffoundation.nl/smartsite.dws?ch=ENG&cid=13180> consultato in data 2008-06-11.
- [19] <epc.ub.uu.se/ETD2007/programme.html> consultato in data 2008-06-11.
- [20] <www.dart-europe.eu> consultato in data 2008-06-11.
- [21] «La tesi di laurea costituisce un'opera dell'ingegno, come tale tutelabile in base alla normativa sul diritto d'autore, in quanto le attività di controllo, sorveglianza ed ingerenza, talvolta anche pregnanti, esercitate dal professore-relatore, non impediscono che la tesi sia il risultato precipuo dell'attività creativa del laureando. Sulla base di tali premesse, non si può certamente negare alla tesi di laurea il carattere di opera creativa, anche qualora il laureando utilizzi argomentazioni e temi già svolti da altri autori o si avvalga di consigli o delle ingerenze del professore-relatore, in quanto, anche in tal caso è individuabile almeno un "minimum" di elaborazione ed organizzazione personale dello studente considerato sufficiente ai fini della tutelabilità del diritto d'autore». Va sottolineato che la sentenza riguardava le tesi di laurea, ma a maggior ragione si è pensato potesse essere estesa, per analogia, alle tesi di dottorato.
- [22] Non ci riferiamo qui a ricerche finanziate da terzi per le quali, di solito, esistono accordi particolari con il soggetto che finanzia la ricerca. Fra le azioni che richiedono autorizzazione dall'avente diritto citiamo la riproduzione e la diffusione, la memorizzazione, l'adattamento e la trasformazione del file depositato ai fini della conservazione.
- [23] L'embargo è un periodo di tempo in cui la tesi archiviata in un deposito istituzionale risulta consultabile solo per la parte relativa ai metadati, non per il *full-text*; il periodo consigliato non oltrepassa i 12 mesi, è però possibile che si arrivi a richiedere fino a un perio-

- do di tre anni.
- [24] Si ricorda tuttavia che l'eventuale articolo o monografia derivanti dalla tesi di dottorato avranno uno stile espositivo e dunque una forma differenti, e saranno quindi da considerarsi una nuova opera rispetto alla tesi di dottorato.
- [25] Tesi di dottorato e diritto d'autore. Indicazioni per l'applicazione delle Linee guida per l'accesso aperto alle tesi di dottorato.
- [26] È consentita la libera pubblicazione attraverso la rete Internet, a titolo gratuito, di immagini e musiche a bassa risoluzione o degradate, per uso didattico o scientifico e solo nel caso in cui tale utilizzo non sia a scopo di lucr.
- [27] Avviato nel 2005, il progetto ha lo scopo di assicurare alle risorse digitali depositate per il lungo periodo la vitalità, traducibilità, autenticità e fruibilità. Cfr. la documentazione del progetto al link: <www.rinascimento-digitale.it/index.php?SEZ=28> consultato in data 2008-06-13.
- [28] Per un approfondimento sul OAI-PMH <www.openarchives.org/pmh> consultato in data 2008-06-15. Il protocollo OAI supporta il Dublin Core semplice.
- [29] Per una descrizione approfondita della procedura di deposito legale cfr. Giovanni Bergamin, *Il deposito per il lungo periodo delle tesi di dottorato in formato digitale: molti problemi si possono risolvere* in Serena Sangiorgi, Eds. *Proceedings Cristalli di esperienza: nuove prospettive e scenari per le tesi di dottorato: conservazione, accessibilità, certificazione, formati, integrazione con Open Access*. "Quaderno CNBA" (10), pp. 21-26, Torino (Italy). <digital.casalini.it/editori/default.asp?codice_opera=19144&numero=8&articolo=3&tipologia=R#> consultato in data 2008-06-15
- [30] Digital Preservation Coalition *Preserving the data explosion: using PDF* reperibile all'URL: <www.dpconline.org/docs/reports/dpctw08-02.pdf> consultato in data 2008-06-13.
- [31] L'archivio è visibile al seguente indirizzo: <amsdottorato.cib.unibo.it> consultato in data 2008-06-13. Per una presentazione approfondita dell'archivio si veda Marialaura Vignocchi, Simone Sacchi, Roberta Lauriola, Enio Carboni, *Le tesi vanno sul web: l'esperienza dell'Università di Bologna* in Serena Sangiorgi, Eds. *Proceedings Cristalli di esperienza: nuove prospettive e scenari per le tesi di dottorato: conservazione, accessibilità, certificazione, formati, integrazione con Open Access*. "Quaderno CNBA" (10), pp. 63-73, Torino (Italy). <digital.casalini.it/editori/default.asp?codice_opera=19144&numero=8&articolo=7&tipologia=R> consultato in data 2008-06-15
- [32] <www.miur.it/0002Univer/0052Cooper/0069Ricono/0356Docume/2257Conven_cf2.htm> consultato in data 2008-06-13.