

Digitalia

Anno VI, Numero 2 - **2011**

Rivista del digitale nei beni culturali

ICCU-ROMA

Il deposito legale delle tesi di dottorato di ricerca: l'esperienza dell'Università di Bologna

Marialaura Vignocchi – Roberta Lauriola
Alma Mater Studiorum, Università di Bologna

Il deposito legale delle tesi di dottorato di ricerca mediante la procedura automatizzata predisposta dalle Biblioteche nazionali centrali di Roma (BNCR) e Firenze (BNCF)¹ è ormai una realtà operativa da diversi anni presso l'Università di Bologna. Ripercorrere oggi le tappe dell'evoluzione dell'iniziativa tracciandone un bilancio è un'occasione per riflettere sui risultati raggiunti e le criticità ancora da affrontare.

A distanza di qualche anno dall'avvio dell'iniziativa possiamo affermare che gli aspetti tecnici e tecnologici del servizio hanno avuto uno sviluppo abbastanza lineare e sono stati tutto sommato meno problematici degli aspetti organizzativi e gestionali che ancora oggi rischiano di limitare la portata della realizzazione degli obiettivi di servizio del progetto complessivo.

Il deposito legale via *harvesting* automatico dei file delle tesi costituisce infatti uno dei servizi a valore aggiunto dell'*institutional repository* AMS Tesi di Dottorato dell'Università di Bologna² che ha come finalità generali:

- migliorare i servizi bibliografici realizzati intorno alle tesi di dottorato dell'Ateneo favorendone la visibilità e la fruizione e promuovendone la loro pubblicazione in rete ad accesso aperto;
- sensibilizzare e abituare i giovani ricercatori all'utilizzo di nuove opportunità di disseminazione dei risultati della ricerca e ad una gestione più consapevole del diritto d'autore;
- razionalizzare i processi di gestione, conservazione e fruizione delle tesi di dottorato di ricerca attraverso la predisposizione di un'unica infrastruttura tecnologica e organizzativa centralizzata.

L'opportunità di semplificazione e razionalizzazione delle procedure amministrative offerta dal deposito legale automatizzato hanno fatto da volano al progetto complessivo. Ad oggi il *repository* istituzionale AMS Tesi di Dottorato, attraverso il quale viene realizzato il deposito legale, ospita tutte le tesi di dottorato di ricerca

¹ <http://www.depositolegale.it/> (15.12.2011)

² <http://amsdottorato.cib.unibo.it/> (15.12.2011)

Progetti

discusse presso l'Ateneo Bolognese a partire dal 2007. Tuttavia, se l'obbligo amministrativo-legale può dirsi pienamente raggiunto con un tasso di tesi auto-archivate da parte dei dottorandi che copre ormai l'intera produzione annua, le finalità generali del progetto - e a ben guardare anche dell'istituto del deposito legale - sono purtroppo ancora lontane dall'essere conseguite in modo soddisfacente.

Tappe della sperimentazione

L'*expertise* tecnica e le infrastrutture sviluppate nell'ambito del progetto di Biblioteca digitale dell'Università di Bologna (AlmaDL)³, unitamente ai rapporti di collaborazione con la BNCf, avviati nel 2004 con la sottoscrizione della convenzione per il deposito volontario dei file⁴, hanno consentito all'Ateneo bolognese di partecipare fin dal principio alla sperimentazione della nuova procedura automatica di deposito legale delle tesi promossa dal GdL OA della Conferenza dei rettori delle università italiane (CRUI)⁵ e dalla BNCf come correttivo all'indicazione di deposito dei CD-ROM e dei DVD prevista dalla circolare MIUR del luglio 2006⁶. Nella primavera del 2007, AlmaDL in collaborazione con l'Ufficio Dottorato di Ateneo, ha predisposto l'archivio aperto istituzionale dedicato e ha avviato la prima campagna sperimentale di raccolta delle tesi in formato digitale - che all'epoca prevedeva ancora la consegna del file su supporto fisico, come da circolare, accompagnato da una liberatoria sottoscritta dal dottorando. Nell'autunno del 2007 lo staff AlmaDL ha caricato nell'archivio AMS Tesi di Dottorato le prime tesi di dottorato di ricerca con cui la BNCf ha potuto testare la fattibilità tecnica della cattura dei metadati e dei file e con cui ha cominciato a mettere a punto le varie fasi del *workflow* di archiviazione permanente.

L'avvallo da parte del MIUR della nuova procedura sperimentale di deposito delle tesi in formato digitale attraverso *harvesting* automatizzato, nel luglio del 2007⁷ e le Linee Guida del GdL OA CRUI⁸ approvate dalla Conferenza nel novembre del 2007, hanno permesso all'Università di Bologna di introdurre l'obbligo di consegna della tesi in formato digitale senza dover attendere il completamento dell'iter istituzionale che, com'è noto, richiede tempi lunghi. Solo nel 2008 si è potuto introdurre l'obbligo di deposito digitale degli elaborati finali nei bandi per i posti di dottorato di ricerca, mentre le finalità complessive del progetto elaborate secondo le raccomandazioni della CRUI, sono state approvate dalla Commissione Ricerca nel 2009 e solo successivamente recepite nel nuovo Regolamento di Ateneo per il Dottorato di Ricerca emanato nell'aprile del 2010⁹.

³ <http://almadl.cib.unibo.it/> (15.12.2011)

⁴ <http://almadl.cib.unibo.it/servizi/deposito-volontario/convenzionebncf.pdf> (15.12.2011)

⁵ <http://www.cruir.it/HomePage.aspx?ref=894> (15.12.2011).

⁶ Circ. MIUR 1420/2006.

⁷ Circ. MIUR 1746/2007.

⁸ CRUI. Gruppo di lavoro OA, *Linee guida per il deposito delle tesi di dottorato negli archivi aperti*: <http://www.cruir.it/homepage.aspx?ref=1149> (15.12.2011).

⁹ Alma Mater Studiorum, Università di Bologna, *Regolamento in materia di dottorato di ricerca*.

Progetti

L'attivazione, in via sperimentale nel 2009 poi a regime nel 2010, di una procedura web di caricamento delle tesi ad uso dei dottorandi ha costituito un notevole miglioramento nella gestione del servizio, consentendoci una riduzione del carico di lavoro in capo allo staff che gestisce l'archivio e un'ottimale e tempestiva copertura dell'obbligo di deposito. Oggi saremmo in grado di "pubblicare" le tesi e renderle disponibili per il deposito legale immediatamente dopo la discussione, se non fosse per i tempi necessari alla registrazione dei verbali degli esiti degli esami finali a carico dell'Ufficio amministrativo. L'archivio oggi (18 ottobre 2011) pubblica 2.896 tesi di dottorato di cui 561 discusse nel corso del 2011.

Aspetti tecnologici: il software e le personalizzazioni

La procedura automatizzata di deposito legale delle tesi di dottorato non può prescindere dalla loro archiviazione in formato digitale all'interno di *repository* che siano compatibili con il protocollo di interoperabilità OAI-PMH¹⁰ così come implementato dai principali software per la realizzazione di archivi aperti. La tecnologia più diffusa è per lo più disponibile *open source*. Si tratta di soluzioni ormai conosciute e di facile implementazione, costantemente aggiornate da *communities* di utenti estremamente attive, che ne garantiscono la manutenzione e lo sviluppo, introducendo ad ogni *release* migliorie e nuove *features* sempre più funzionali alle esigenze degli utenti. Il livello di evoluzione e la frequenza di aggiornamento di applicativi come EPrints¹¹ e DSpace¹² è tale che non sembra consigliabile affidarsi a tecnologia proprietaria e sviluppata ad hoc sia per i costi elevati sia per il rischio di rapida obsolescenza delle soluzioni adottate.

La tecnologia scelta per AMS Tesi di Dottorato è GNU EPrints. Le personalizzazioni a questo sw sono state realizzate più per esigenze organizzative interne che non per la procedura di deposito legale. Le personalizzazioni hanno riguardato i seguenti aspetti:

- i set dei metadati
- il sistema di autenticazione e di profilazione degli utenti
- il flusso di lavoro della fase di deposito
- la gestione dei profili del diritto d'autore

Sono stati individuati due set di metadati, mappati uno sull'altro. Il primo, più esteso, è il set utilizzato internamente da EPrints per la descrizione delle tesi che consente di offrire "viste" personalizzate per denominazione del Dottorato di ricerca, Scuola di dottorato, settore scientifico disciplinare MIUR etc.

http://www.normateneo.unibo.it/NormAteneo/Regolamento_in_materia_di_Dottorato_di_Ricerca.htm (15.12.2011).

¹⁰ Open Archives Initiative Protocol for Metadata Harvesting: <http://www.openarchives.org/pmh/>(16.12.2011).

¹¹ <http://www.eprints.org/>(16.12.2011).

¹² <http://www.dspace.org/> (16.12.2011).

Progetti

Il secondo è costituito dalla mappatura Dublin Core che permette il colloquio con la BNCF tramite interfaccia OAI-PMH. Per quest'ultimo set si sono seguite le indicazioni pubblicate sul sito depositolegale.it¹³ che hanno recepito le raccomandazioni delle *Linee Guida* CRUI¹⁴ basate, a loro volta, su iniziative internazionali come *Knowledge Exchange*¹⁵ e *Driver*¹⁶.

L'analisi dei metadati ha permesso di individuare la presenza di database di Ateneo che potessero essere integrati per garantire consistenza/coerenza dei dati inseriti nel deposito. L'architettura a *plug-in* di EPrints ha molto facilitato l'integrazione di componenti aggiuntivi nel software per il colloquio con database esterni, senza intaccarne la struttura portante, e garantendo, in caso di aggiornamento, una buona compatibilità. In particolare è stato sviluppato un modulo di autenticazione che utilizza il *webservice* DSA Authentication di Ateneo per la registrazione, la profilazione e l'autenticazione dei dottorandi chiamati a caricare il file della loro tesi. Il modulo cattura inoltre i dati personali e di afferenza istituzionale dell'utente, nell'ottica di limitare al massimo i campi da compilare a cura del dottorando.

A livello di *workflow*, il progetto ha previsto di distribuire sui dottorandi parte del carico di lavoro, attraverso l'**auto-archiviazione** delle tesi in formato elettronico nell'archivio che permette non solo di sollevare l'Ufficio dalla gestione delle copie su supporto cartaceo e/o ottico, ma anche di pre-catalogare le tesi legandole ai dati dell'autore, ricavati dai sistemi informativi di ateneo. Ai dottorandi è stata quindi fornita un'interfaccia per il caricamento del file e dei metadati descrittivi che i pochi contatti ricevuti dal servizio di supporto confermano essere semplice ed amichevole. Il *workflow* prevede, prima della pubblicazione, un ultimo step di controllo da parte dello staff di AlmaDL a garanzia della qualità e correttezza formali dei dati e dei file che vengono esposti tramite l'archivio.

Altro aspetto importante che ha richiesto una personalizzazione è stata l'esigenza di disciplinare i profili del diritto d'autore e di ottenere l'autorizzazione per il trattamento della tesi nell'archivio in attesa delle modifiche ai regolamenti e ai bandi. Per questo motivo la scelta dei diversi livelli di accesso è stata collegata ad altrettanti *disclaimer* legali che vanno a completare, assieme ai dati personali dell'autore e descrittivi della tesi, un documento di liberatoria da stampare al termine dell'auto-archiviazione. In attesa di una migliore integrazione tra i sistemi informativi esistenti tale documento da consegnarsi a cura del dottorando all'Ufficio amministrativo costituisce prova dell'avvenuto deposito della tesi in formato digitale.

¹³ <http://www.depositolegale.it/> /(16.12.2011).

¹⁴ Op. cit.

¹⁵ <http://www.knowledge-exchange.info/> /(16.12.2011).

¹⁶ http://www.driver-support.eu/documents/DRIVER_Guidelines_v2_Final_2008-11-13.pdf/ (16.12.2011).

Progetti

Ai fini del deposito legale, una volta mappato il set Dublin Core secondo le indicazioni previste, gli interventi tecnici si sono limitati alla configurazione del formato standard MPEG21 DIDL¹⁷, per una corretta esposizione dei metadati strutturali in caso di tesi composte da più di un file e all'abilitazione dell'indirizzo IP dell'*harvester* della BNCf per consentire la cattura delle tesi soggette ad embargo.

Aspetti organizzativi e gestionali

La raccolta delle tesi di dottorato di ricerca in formato digitale ha avuto un certo impatto organizzativo soprattutto sullo staff di AlmaDL, che ne gestisce il servizio di assistenza rivolto agli utenti, si fa carico del controllo dei dati inseriti e della pubblicazione delle tesi. I primi anni, in particolare, sono stati critici in quanto la raccolta dei file tramite CD-ROM e DVD ha fatto registrare una serie di problemi imprevisti. L'impossibilità di controllare il contenuto dei dischi alla consegna si è tradotta in un'interminabile rincorsa ai dottori di ricerca che avevano consegnato file protetti, corrotti, mutili, spezzettati in capitoli o con allegati di vario tipo. La gestione dei dischi a distanza di tempo dalla discussione ha reso questa operazione di recupero estremamente difficoltosa. L'esperienza per quanto onerosa è stata tuttavia formativa poiché ha permesso di focalizzare le criticità connesse al processo e i necessari correttivi.

Grazie a questa prima fase è stato possibile fornire agli studenti - che dal 2009 hanno l'obbligo di auto-archiviare la tesi direttamente in AMS Tesi di Dottorato - istruzioni precise sul formato, il nome, i contenuti del file e sull'inserimento dei pochi metadati descrittivi non derivati. Il controllo periodico dei file caricati nell'archivio, specialmente in prossimità delle due scadenze per la consegna degli elaborati finali, permette, in caso di problemi, di intercettare gli studenti in tempo utile prima della discussione della tesi. I problemi più frequenti sono rappresentati dalla mancanza del frontespizio o dall'errato deposito di diverse versioni della tesi da parte dello stesso studente. A volte occorre intervenire sui contenuti per ragioni connesse alla tutela del diritto d'autore, quando ad esempio, incorporati nella tesi si trovano i testi integrali di pubblicazioni tutelate, sia dell'autore che di terzi. A questi inconvenienti si può far fronte con una corretta campagna di informazioni preventiva che dovrebbe coinvolgere non solo i dottorandi, ma i docenti che svolgono funzioni di tutor nei dottorati di ricerca.

Al "validatore" compete comunque il controllo finale, precedente la pubblicazione, dei file e dei metadati perché possano essere correttamente esposti attraverso l'archivio, anche ai fini del deposito legale. In particolare, si tratta di correggere eventuali refusi nel titolo o di normalizzare l'uso del maiuscolo, ammesso solo dove ortograficamente richiesto; spesso occorre intervenire sui caratteri speciali dei testi incollati dagli utenti nel campo abstract.

¹⁷ <http://mpeg.chiariglione.org/standards/mpeg-21/mpeg-21.htm> /(16.12.2011).

Progetti

È auspicabile, inoltre, un intervento professionale più qualificato che preveda l'inserimento di descrittori semantici di soggetto o di classe, ma al momento, purtroppo, risulta troppo oneroso.

Altre attività che impattano sullo staff che gestisce l'archivio sono le richieste di sostituzione del file della tesi, sia prima che dopo la discussione, le richieste di modifica del livello di accesso successivamente alla discussione e la certificazione di avvenuto deposito legale. Per quanto riguarda le esigenze di modifica dei file e del livello di accesso dopo la discussione della tesi si è pensato di regolamentare il servizio per impedire che tali richieste possano essere avanzate dal momento che non paiono legittime. Vengono invece accettate le richieste di sostituzione dei file prima dell'esame finale poiché quest'ultima prassi appare legata a rigidità di tipo amministrativo – la discussione avviene molto tempo dopo la data di consegna fissata dagli uffici. Forse nell'ambito di un processo interamente automatizzato, tali vincoli potrebbero essere superati.

La certificazione di avvenuto deposito viene eseguita sulla base dell'evidenza che il servizio di deposito invia successivamente alle attività di *harvesting*. Si tratta di una comunicazione via e-mail con l'identificativo (URI) della tesi e l'impronta digitale in formato SHA-1 base 32. La mail viene inviata all'account di posta che risulta inserito come amministratore dell'archivio e che viene esposto tra i metadati attraverso il server OAI. In aggiunta, il servizio di deposito legale fornisce a tutte le università che aderiscono alla procedura un account che consente di accedere all'archivio di conservazione e verificare direttamente lo stato di un preciso documento di cui si conosce l'URL.

La questione dell'accesso

Al di là degli aspetti tecnici, amministrativi e organizzativi, il cuore vero della questione rimane il problema della fruibilità del testo pieno delle tesi di dottorato. La finalità ultima sia dei *repository* istituzionali sia del deposito legale è in fin dei conti consentire la scoperta e la consultazione di documenti che spesso racchiudono l'eccellenza della ricerca scientifica dei nostri atenei e che non sempre riescono, in tempi brevi, a prendere la forma definitiva di pubblicazioni ufficiali: articoli o libri. Per questo, l'adozione di una *policy* istituzionale per la fruibilità del testo pieno delle tesi di dottorato di ricerca che ne favorisca la loro pubblicazione in rete ad accesso aperto, assieme ad una corretta gestione del diritto d'autore e a una diffusa informazione ai docenti e ai dottorandi relativa alle finalità del servizio costituiscono aspetti cruciali per la realizzazione degli obiettivi di servizio.

L'introduzione dell'obbligo amministrativo in ottemperanza alla norma sul deposito legale non basta a risolvere il problema dell'accesso e della fruibilità del testo. Anzi la maggiore e più puntuale disponibilità in rete dei metadati delle tesi di dottorato di ricerca, indicizzate dai motori di internet e dai servizi dedi-

Progetti

cati,¹⁸ rischia di generare un flusso di richieste informative a cui è impossibile rispondere in assenza di una precisa politica dell'accesso. Per questi motivi l'attivazione della procedura di auto-archiviazione delle tesi di dottorato da parte degli stessi dottorandi, entrata pienamente a regime nel 2010, ha generato un imprevisto effetto collaterale che abbiamo visto pericolosamente stabilizzarsi nel 2011. Nei primi due anni di vita del servizio la consegna della tesi avveniva su DVD o CD ed era mediata dal personale d'ufficio che aveva modo di spiegare lo scopo e la portata dell'iniziativa, ricordando agli studenti le indicazioni della CRUI. Lo studente così informato, nel sottoscrivere la liberatoria, sceglieva quasi sempre l'accesso aperto o un embargo ragionevole intorno ai 12 mesi. Con l'attivazione della procedura su web, la maggior parte dei dottorandi, lasciati soli nella decisione dei livelli di accesso, ha optato per quelli più restrittivi, "chiudendo" la tesi per sempre o scegliendo periodi di embargo lunghissimi. Senza un'adeguata campagna di informazione e senza l'adozione di una *policy* dell'accesso, un'implementazione tecnica che ha reso il servizio più efficiente e sostenibile, ne sta limitando l'efficacia.

Occorre dunque definire chiaramente le modalità che consentono di fruire dei contenuti informativi delle tesi di dottorato e non solo rendere disponibili le informazioni bibliografiche o effettuare il deposito legale. È necessario, inoltre, ridefinire i servizi bibliografici che gli istituti deputati al deposito e alla conservazione nel tempo possono offrire perché la conoscenza possa circolare, sfruttando pienamente le possibilità offerte dal digitale e dalle reti. È verosimile che le soluzioni migliori e più facilmente sostenibili debbano tenere in conto, bilanciandoli, gli interessi e i diritti di tutte le parti in gioco: autori, università, biblioteche e lettori.

¹⁸ Le tesi di dottorato dell'Università di Bologna sono indicizzate da una serie di servizi fra cui spiccano: Dart-Europe E-Theses Portal, <http://www.dart-europe.eu/basic-search.php>; NDLTD, <http://www.ndltd.org/>; driver, <http://www.driver-support.eu/index.html>; WorldCat, <http://www.worldcat.org/> (tutti i siti sono stati consultati il 16.12.2011).