

L'utilizzo dell'archivio istituzionale come strumento di knowledge management dell'istituzione universitaria

Paola Galimberti (Unimi) Alberto Silvani (Unimitt)

Abstract

Gli archivi istituzionali nascono come strumenti di raccolta organizzata delle produzioni scientifiche delle istituzioni universitarie. In questo senso sono stati considerati fin dal loro avvio strumenti di pertinenza dei bibliotecari i quali hanno dato un prezioso contributo al loro sviluppo. Recentemente ha però preso avvio una seconda fase nell'evoluzione degli archivi, che ne ha individuato le potenzialità "informative" e che li pone al centro di una rete che collega l'autore, il gruppo di ricerca e il dipartimento con i servizi istituzionali al fine di una più efficiente e globale gestione del sistema ricerca dell'ateneo. La presente relazione analizza alla luce dell'esperienza dell'Università degli Studi di Milano, la crescente domanda di servizi e informazioni a valore aggiunto posta all'archivio e le soluzioni implementate nell'ottica della maggiore integrazione possibile fra fonti informative interne ed esterne.

Keywords: LIS, Knowledge Management, Research Assessment, Information Resources, Institutional repositories

1. Il contesto e le problematiche.

Gli archivi istituzionali sono collezioni digitali dei lavori prodotti all'interno di una istituzione di ricerca. Lo scopo principale per cui sono creati è quello di documentare in maniera permanente ed esaustiva, in un unico punto, l'attività di ricerca svolta all'interno dell'ateneo. Essi fungono da indicatori tangibili della qualità di una istituzione aumentandone la visibilità, il prestigio e il valore pubblico¹. Gli archivi contengono dati bibliografici, amministrativi e i documenti di ricerca stessi a cui è garantito il libero accesso da parte di chiunque sia dotato di un computer e di una connessione. Sono circa 1400 gli archivi di questo tipo nel mondo (di cui 47 in Italia)².

Questi strumenti sono organizzati sulla base di requisiti standard (vale a dire che tutti gli archivi, indipendentemente dal software utilizzato, espongono i metadati dei lavori contenuti nella stessa modalità). Il protocollo che utilizzano è l'Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH³), un protocollo che permette l'interoperabilità, la conservazione a lungo

¹ C. Lynch, (2003) sottolinea come i contenuti digitali in un sistema non coordinato siano a rischio. Nelle istituzioni vi è una grande massa di informazioni digitali decentrate che devono essere in qualche modo raccolte, condivise e preservate dall'organizzazione. L'archivio istituzionale nasce anche con questo fine

² Per quanto riguarda il numero di documenti contenuti negli archivi, Oaister indicizza 22,742,966 documenti (dati agosto 2009)

³ <http://www.openarchives.org/OAI/openarchivesprotocol.html>

termine e l'indicizzazione da parte dei maggiori motori di ricerca generalisti e specialistici (Google, Yahoo, Google Scholar) e di quelli specifici del mondo Open Access (Oaister⁴, Scientific Commons⁵ ecc.)

Poiché, come sottolineato, accanto ai metadati bibliografici e amministrativi, gli archivi contengono anche i testi pieni dei lavori di ricerca permettendone la libera consultazione, essi forniscono uno strumento formidabile per una comunicazione immediata dello stato dell'arte della ricerca in un determinato ambito disciplinare o territoriale, funzione che può portare ad evitare la duplicazione dei progetti, o condurre a nuove collaborazioni fra discipline diverse sia all'interno che all'esterno dell'istituzione e rappresentare l'occasione per facilitare le relazioni tra il mondo della ricerca e i suoi interlocutori "esterni".

Proprio per le loro caratteristiche di "contenitori" di documenti e dati bibliografici questi strumenti sono stati considerati fin dalla nascita di stretta pertinenza delle biblioteche. E in effetti molti dei servizi dedicati ad autori, contributori ed utenti sono di ambito squisitamente bibliotecario: conservazione a lungo termine, migrazione dei contenuti, disseminazione e gestione dell'accesso, gestione di metadati e di formati, motori di ricerca, data mining ecc. Di fatto i bibliotecari hanno contribuito non poco al loro sviluppo e alla loro manutenzione, verificando e validando costantemente i dati in entrata, migliorandone l'architettura, modellando il flusso di validazione, collaborando con gli informatici per un funzionamento più efficiente. Non è dunque raro che gli archivi siano gestiti dai sistemi bibliotecari⁶. In questo contesto si è sottolineata e curata soprattutto la funzione di catalogazione, organizzazione ed esposizione dei dati bibliografici e quella di ripubblicazione (compatibilmente con la legge sul diritto d'autore) dei lavori di ricerca già certificati dagli editori.

Abbastanza presto però, anche altre strutture delle istituzioni hanno cominciato a capire come all'interno degli archivi si nascondano una serie di dati amministrativi e gestionali significativi per una più completa rappresentazione (e valutazione) dell'attività di ricerca svolta nell'ateneo. Nel corso della loro pur breve vita essi sono stati così chiamati/piegati ad offrire servizi sempre più raffinati e differenziati a soggetti diversi all'interno delle istituzioni. Si è cominciato ad estrarre i dati contenuti pensando a finalità diverse dalla mera documentazione dell'attività svolta dal singolo ricercatore: centralmente per la valutazione o la valorizzazione verso terzi, a livello di utente/ricercatore per la creazione di liste di pubblicazioni per i CV, per le relazioni sull'attività dei Dipartimenti, come supporto per la selezione delle pubblicazioni per gli esercizi di valutazione a livello nazionale, per la creazione di report su gruppi di ricerca e progetti. Il tutto è andato a

⁴ <http://www.oaister.org/>

⁵ <http://www.scientificcommons.org/>

⁶ Un recente rapporto della Association of Research Library (2009, p. 9) individua nella gestione dell'archivio istituzionale la funzione cruciale di una biblioteca di ricerca.

sommarsi ad impieghi già esistenti e, soprattutto, ad architetture già impostate e quindi da reimpostare e rimettere a punto ogni volta.

Poiché la domanda di servizi a valore aggiunto cresce, diventa necessario riconsiderare il modo in cui l'archivio istituzionale può contribuire allo sviluppo strategico-gestionale dell'istituzione.

Può essere utile guardare a questo passo successivo in termini di gestione (o meglio capacità di gestione) dei contenuti di conoscenza e dell'organizzazione. In altri termini di knowledge management.

2. La domanda crescente di servizi a valore aggiunto (richieste e nuove professionalità)

Le università sono per mission centri di produzione di conoscenza, una conoscenza ad alto livello e che assume le forme più svariate: brevetti, marchi modelli, pubblicazioni, esperienza personale dei singoli ricercatori, processi, pratiche, norme ecc. Le stesse sono sollecitate a stabilire relazioni e a rendere disponibili le loro conoscenze anche per acquisire quelle risorse che faticano a trovare all'interno del tradizionale circuito accademico. A questo si sommano le nuove modalità di produzione di conoscenza che si basano su interazioni e appartenenze a reti e comunità di operatori attivi in diversi ambiti. In questo contesto l'archivio, nel divenire – almeno potenzialmente – uno strumento utilissimo per diversi scopi, manifesta anche tutti i propri limiti a svolgere pienamente una funzione che non sia di mero supporto al flusso delle informazioni necessarie alle decisioni, al trasferimento e alla creazione di conoscenza.

In un futuro non troppo lontano l'archivio potrebbe rappresentare un supporto per le “comunità di pratiche” dove la conoscenza è creata e condivisa in un ambiente affidabile. Se, da una parte, la conoscenza di ciò che stanno facendo altri gruppi di ricerca nell'istituzione stessa (conoscenza che ora non è affatto scontata in istituzioni grandi e multidisciplinari) può portare alla creazione di nuove comunità interdisciplinari, alla valorizzazione delle esperienze interne anche in settori disciplinari molto diversi e soprattutto alla collaborazione con soggetti terzi anche commerciali, la presenza, dall'altra, dei risultati di ricerca ad accesso aperto crea una vetrina della produzione scientifica dell'università che, nell'essere liberamente interpretata e “valorizzata” dagli utilizzatori, permette di sottolineare quali sono le aree strategiche della ricerca svolta e gli attori principali che possono condurre alla formazione di nuovi gruppi di eccellenza.

In un contesto così delineato l'archivio è anche fattore di cambiamento culturale. Da un lato nel fornire una vetrina delle attività (e quindi costringendo a rendere pienamente visibile l'insieme dell'output degli individui e delle strutture); dall'altro, nel consentire esercizi collegati alla valutazione, sia da parte delle strutture deputate sia da parte di chiunque sia in grado di operare sui dati a disposizione, contribuendo con ciò al processo di qualificazione e validazione dei dati inseriti, che provengono, nella maggioranza dei casi, da un'alimentazione individuale di matrice bottom-up.

La necessità di valutare a vari livelli, ha condotto, a seconda degli obiettivi, alla richiesta di report sempre più specifici e a diversa granularità sulle attività di ricerca di gruppo, progetto, centro di ricerca, dipartimento; rapporti che integrino le informazioni relative alla produzione scientifica, ai finanziamenti, alla rendicontazione, all'anagrafica. In questo senso l'archivio entra, come tassello fondamentale, a fare parte di un sistema che mira a integrare set di informazioni provenienti da

banche dati diverse sia interne che esterne all'ateneo. E' dunque auspicabile che esso sia collegato all'anagrafica, alla banca dati che gestisce il finanziamento dei progetti, alla contabilità⁷. Questa integrazione permette una descrizione dell'andamento della ricerca di ateneo e una gestione dell'attività di ricerca con un occhio all'ottimizzazione di finanziamenti e costi.

Poiché l'archivio contiene informazioni sensibili, in molti casi oggetto di meccanismi di tutela o comunque contrattualizzate (si pensi a quelle relative ad attività commissionate o sponsorizzate da terzi), in un ambiente affidabile è importante che a tutti i livelli, sia di ufficio legale, che di singolo ricercatore, si dia importanza e ci sia consapevolezza delle questioni legate ai diritti di proprietà intellettuale, dei vincoli cui si è scelto di sottostare, delle regole che si sono accettate sottoscrivendo i contratti. Il supporto legale alle strutture di ricerca è fondamentale per una istituzione che si prefigga una gestione globale del sistema ricerca.

Non si tratta solo di sapere interpretare le policy degli editori rispetto alla ripubblicazione dei testi nell'archivio (quale versione pubblicare, dove pubblicarla, con quali vincoli ecc.) ma di affrontare questioni riguardanti la compensazione dei diritti di terzi, i rischi relativi a future pubblicazioni, la gestione degli embarghi per salvaguardare materiali potenzialmente brevettabili o comunque sfruttabili commercialmente e le questioni relative alla privacy. Anche per questo è necessario sviluppare competenze che non sono più squisitamente di ambito bibliotecario, ma che riguardano la gestione dei contratti di edizione, la normativa brevettuale, la conoscenza e interpretazione delle policy editoriali, la normativa sulla privacy.

Se inquadrriamo l'archivio istituzionale nel contesto dei processi di Knowledge Management (KM) è possibile identificare alcuni dei servizi necessari per il suo successo. Allo stato attuale gli archivi sono nella fase di sviluppo e di crescita del loro ruolo all'interno delle istituzioni e necessitano di un sostanzioso contributo nel delineare la loro funzione nei processi della conoscenza istituzionale. Il KM può fornire un significativo aiuto in questa direzione.

3. L'esperienza all'interno dell'Università degli studi di Milano

L'Archivio Istituzionale della Ricerca (AIR⁸) nasce come strumento di supporto alla produzione della relazione annuale dei Dipartimenti. All'Ateneo mancava un punto di raccolta centrale dell'attività di ricerca svolta nell'istituzione, così che frequentemente non si veniva a conoscenza e non si dava visibilità a produzioni importanti per la valorizzazione della ricerca verso l'esterno (sia in termini di brevetti che in termini di pubblicazioni scientifiche) e verso l'interno. In questo senso la costituzione di un "magazzino/vetrina" dell'output rappresenta un passaggio di razionalità nella gestione delle informazioni utili all'Amministrazione.

Alcuni dati di AIR sono precaricati (dati anagrafici, afferenza, settore scientifico disciplinare, area scientifica, facoltà), altri vengono inseriti dai diretti interessati. Solo chi fa ricerca ed è strutturato ha la possibilità caricare i dati per cui per accedere al sistema è necessario autenticarsi. Alla fine dell'immissione è richiesta la sottoscrizione di un licenza in cui si danno una

⁷ Sistemi di Ateneo

⁸ <http://air.unimi.it>

serie di garanzie sulla veridicità dei dati, sulla compensazione dei diritti, sulla paternità dell'opera e si autorizza l'Ateneo al trattamento dei dati e dei files a fini amministrativi e di conservazione. L'archivio è alimentato da oltre 2700 fra docenti e ricercatori. Poiché manca (ancora) una consapevolezza diffusa dell'importanza della qualità dei dati immessi e nonostante linee guida e indicazioni presenti nella maschera di inserimento, tutte le registrazioni inserite devono essere riviste e validate dai bibliotecari, ognuno nelle proprie aree di competenza⁹.

Il lavoro di personalizzazione dei metadati (ancora in corso) è stato lungo e complesso, e ha dovuto tenere conto del contesto (e degli usi) della ricerca interna, ma anche di quella internazionale, delle esigenze dei comitati di valutazione, della interoperabilità con archivi bibliografici e amministrativi interni ed esterni. Da subito si è provveduto al collegamento con l'anagrafica dell'Ateneo, dando così ai ricercatori la possibilità di autenticarsi attraverso l'LDAP di ateneo (un protocollo di livello applicativo per interrogare e modificare servizi di directory distribuiti tramite reti TCP/IP). L'acquisizione dei dati dall'anagrafica consente un allineamento costante delle informazioni e l'assegnazione automatica dei ricercatori alla struttura di riferimento, con conseguente profilazione.

L'implementazione dell'archivio ha richiesto l'integrazione e la collaborazione con varie strutture dell'ateneo: l'area del personale e delle segreterie (per la gestione del personale strutturato e non); la divisione sistemi informativi (per l'attivazione dei sistemi di autenticazione e scambio di dati con altre piattaforme dell'ateneo); il nucleo di valutazione (per l'analisi delle procedure di valutazione e la strutturazione dei dati), la divisione servizi per la ricerca (per la gestione dei progetti in ambito nazionale e internazionale), l'ufficio legale e il Centro per il trasferimento tecnologico (per la profilazione dei ricercatori, le questioni relative al diritto d'autore e per le procedure relative al deposito di tesi di dottorato e brevetti)¹⁰.

I direttori di dipartimento hanno cominciato ad utilizzare l'archivio per estrarre i dati sulle pubblicazioni da inserire nella relazione annuale che compilano sulla base di un modello fornito dalla CRUI. L'archivio è stato integrato anche con il modulo per la richiesta di finanziamento dei progetti, da cui possono essere richiamate le pubblicazioni che il sistema prevede debbano essere integrate nella domanda.

In questa prima fase la maggiore criticità è stata rappresentata dalla completezza dell'archivio. Non c'era incentivo ad inserire i dati delle proprie pubblicazioni, poiché comunque gli stessi dati andavano caricati sul sito docente della banca dati Cineca¹¹. Il problema è stato risolto in due step. Da un lato si è fatto in modo che i dati dell'archivio potessero andare ad alimentare

⁹ A titolo esemplificativo il margine di errore, ovvero la quantità di informazioni inserite che necessitano di correzioni sono pari al 50% per errori materiali, al 70% per incompletezza dei dati, mentre rimane elevata la quota di non inserimenti legata al fatto che per molti SSD l'archivio non viene considerata la sede prioritaria per il caricamento dei dati

¹⁰ Recentemente è stato acquistato un sistema per la gestione della ricerca di Ateneo (Surplus) che integra diversi moduli fra cui l'archivio istituzionale.

¹¹ Banca dati del MIUR

quotidianamente il sito docente CINECA (ma non vice-versa) evitando così il doppio inserimento ma privilegiando la direzionalità, dall'altro una delibera del Senato ha reso obbligatorio il caricamento dei dati riguardanti i lavori di ricerca entro un mese dalla pubblicazione¹². Come tutti gli archivi istituzionali AIR prevede oltre ai metadati anche l'inserimento dei full-text. A tal fine sono state elaborate della FAQ e forniti degli strumenti che servano agli autori da indirizzo rispetto al riutilizzo dei propri testi in maniera legalmente corretta e rispettosa delle norme del copyright e del diritto d'autore. I bibliotecari che nel corso di questi anni si sono fatti carico della validazione dei dati immessi dai ricercatori hanno sempre agito da filtro nel caso di violazioni del diritto d'autore

Nella seconda fase che ha preso avvio di recente, nell'ottica di una migliore e più efficace gestione dei progetti di ricerca, si è collegato l'archivio ad un modulo che gestisce le diverse tipologie di progetto e la loro rendicontazione così da metterlo al centro di una rete che collega l'autore, il gruppo di ricerca e il dipartimento con i servizi istituzionali. Si è venuto così a costituire un vero e proprio modulo di "repertorio delle competenze" in grado di integrare l'output con le anagrafiche degli individui, delle strutture e dei progetti e contratti.

Come già ricordato, la presenza di informazioni strutturate secondo standard internazionali e l'accesso diretto ai risultati aprono nuove possibilità alla valutazione della qualità e dell'impatto della produzione scientifica, in particolare per quelle aree che hanno finora registrato le maggiori difficoltà nell'uso degli strumenti bibliometrici tradizionali come per esempio quelle umanistiche e delle scienze sociali. A fronte di una mole centralizzata di dati coerenti e strutturati non è stata però ancora adottata dall'Ateneo una politica articolata su come questi dati possano essere utilizzati e con quali finalità.

Un primo utilizzo, in vista del prossimo esercizio di valutazione quinquennale, potrebbe essere quello di individuare le migliori pubblicazioni sulla base di indicatori quantitativi (il fattore di impatto ad esempio). I migliori risultati dovranno anche essere messi in relazione ai finanziamenti ricevuti e a parità di output si potranno individuare quali aree hanno impiegato i finanziamenti in maniera più efficiente.

L'archivio è però fonte anche di altri indicatori. Ad esempio indicatori sulla produttività (divisa per tipologia) di singoli individui, gruppi di ricerca, dipartimenti. Ciò permette di individuare quali sono le tipologie di prodotti più diffuse nelle varie comunità disciplinari e di confrontare settori scientifico disciplinari afferenti a dipartimenti diversi. Una scarsa produttività in un settore disciplinare che in altri dipartimenti è produttivo può essere indice della cattiva gestione di un dipartimento. Gli stessi dati potrebbero servire anche da orientamento per le scelte delle pubblicazioni da finanziare in futuro a fronte di modelli economici diversi (open access ad es.)

Altri indicatori interessanti che si possono estrarre dall'archivio riguardano il grado di collaborazione con gruppi o ricercatori esterni, o il livello di internazionalizzazione (partecipazione a gruppi di lavoro stranieri, utilizzo della lingua inglese), entrambi utilizzabili per l'analisi di network.

¹² Recentemente il Senato ha ulteriormente rafforzato il ruolo istituzionale di AIR designandolo come fonte prioritaria per gli esercizi di valutazione della ricerca.

Per potere fare questo, però, l'archivio deve sviluppare la sua capacità di gestione delle informazioni contenute anche dialogando con altri archivi: basti pensare alla possibilità di integrarsi con le tradizionali basi dati per le analisi bibliometriche, quali Scopus o Web of Science, ma anche con gli esercizi che sempre più vengono proposti da quelle aree scientifiche che faticano a riconoscersi nella bibliometria tradizionale. Questi esercizi procedono a riclassificare, attraverso l'assegnazione di "giudizi di valore" e con metriche variabili da caso a caso, gli output o, più spesso le tipologie e i "contenitori" attraverso cui gli output vengono veicolati. Non va inoltre dimenticato che, accanto alle varie casistiche di pubblicazioni, un numero sempre maggiore di fonti entrano a far parte del processo di produzione e circolazione di nuove conoscenze, con forme impensabili solo pochi anni fa (ad es. la liquid pub¹³ e tutta l'interazione via web, gli strumenti di social network e social tagging). L'archivio non è ancora lo strumento ottimale ma rappresenta un passo in avanti, a patto che sia in grado di reagire a nuovi stimoli e sia sottoposto ad una gestione dinamica e non conservativa. Ovvero che non resti un'appendice della visione tradizionalmente bibliotecaria e documentativa.

4. Conclusioni

Individuare le conoscenze "nascoste" negli archivi istituzionali di ricerca costituisce un campo di studio relativamente recente, che cresce in proporzione alla loro diffusione e alla possibilità di sperimentazione su realtà diverse ma correlabili. Le tecniche di KM, applicate in maniera non tradizionale, possono offrire un validissimo supporto nell'estrazione delle informazioni e nell'identificazione di legami e relazioni. In questa direzione i contenuti (qualità, completezza ed accessibilità dei metadati) giocano un ruolo fondamentale. E' tuttavia necessaria anche una padronanza – non meramente "difensiva" - degli aspetti legati ai vincoli (e alle opportunità) collegati alla gestione degli aspetti legali (proprietà intellettuale, diritto d'autore, vincoli contrattuali).

Ma le stesse tecniche sono sicuramente utilizzabili per migliorare la gestione degli archivi, troppo spesso considerati come una forma (più moderna ma sostanzialmente in continuità) dei tradizionali contenitori di elenchi di pubblicazioni, destinati ad un uso principalmente bibliotecario. Servono per questo nuove professionalità, o meglio la contaminazione di professionalità esistenti, oltre alla necessità di vedersi riconosciuto un ruolo sulla base dei risultati conseguiti.

Il probabile esercizio di valutazione promosso a livello ministeriale, se da un lato ha visto rivitalizzarsi i soggetti che erogano servizi sul versante della bibliometria, può costituire un interessante momento di verifica delle potenzialità dell'archivio istituzionale, sia sul versante manageriale interno degli atenei nel saper correttamente attingere alle informazioni già possedute, sia nella logica di marketing che inevitabilmente deriva dal posizionamento competitivo nei confronti dei concorrenti. L'esperienza che stiamo maturando all'Università di Milano, pur con le sue luci ed ombre, lascia ben sperare.

¹³ La pubblicazione liquida viene così definita dai suoi teorizzatori "an evolutionary, collaborative, multi-faceted knowledge object that can be composed and consumed at different levels of details" (Casati, Giunchiglia, Marchese, 2007)

Bibliografia:

ARL (Association of Research Libraries), *The Research Library's Role in Digital Repository Services: Final Report of the ARL Digital Repository Issues Task Force*, ARL, 2009
<http://www.arl.org/bm~doc/repository-services-report.pdf>

F. Casati, F. Giunchiglia, M. Marchese, (2007) *Liquid Publications: scientific publications meet the web: changing the way scientific knowledge is produced, disseminated evaluated and consumed*,
<http://eprints.biblio.unitn.it/archive/00001313/01/073.pdf>

M. A. Drake, (2005) *Encyclopedia of library and information Science*, 2nd edition, Taylor and Francis

C. Hess, E. Ostrom, (2009) *La conoscenza come bene comune: dalla teoria alla pratica*, B. Mondadori

C. Lynch, (2003) *Institutional Repositories: Essential Infrastructure For Scholarship In The Digital Age* portal: Libraries and the Academy - Volume 3, Number 2, pp. 327-336

W. White, (2009) *Institutional repositories. Making a contribution to institutional knowledge management and the global research commons*, OR Conference 2009
<http://eprints.soton.ac.uk/48552/>