

Gli archivi e-prints in Italia

di Antonella De Robbio

Responsabile Settore Progetti e Biblioteca Digitale

CAB - Centro di Ateneo per le Biblioteche

Università degli Studi di Padova

coordinatore alla

Tavola rotonda: "Gli archivi e-prints in Italia"

L'Archivio E-Prints dell'Università di Firenze: prospettive locali e nazionali

Firenze, 10 febbraio 2004, Aula Magna, P.za San Marco, 4

*"Le relazioni presentate ad una conferenza sono istantanee
di progetti di ricerca,
presi ad un dato momento dell'anno.
Gli atti di convegno sono un album di istantanee.
I fotografi cercano di fare del loro meglio con il soggetto da presentare
mettendo insieme sufficiente materiale grezzo
e affinando i vari aspetti.
Non essendo consentito il "trucco",
non tutti gli aspetti risultano limati
al momento dello scatto dell'istantanea flash
e perciò qualcosa può non risultare perfetto."*

LUISELLA GOLDSCHMIDT-CLERMONT,

*Modelli di comunicazione nella fisica delle alte energie,
pre-print, febbraio 1965¹*

A qualche settimana di distanza dalla tavola rotonda fiorentina sugli archivi di e-prints in Italia, e nel bel mezzo di un lavoro di traduzione di un vecchio articolo sui pre-print², sto tentando da una parte di mettere ordine agli appunti abbozzati su fogli volanti scarabocchiati, alle annotazioni poste ai margini delle stampate degli interventi, dall'altra di organizzare i pensieri, i memo ancora freschi emersi a seguito della vivace discussione in sede di dibattito, ai tanti discorsi aperti, discussi, frantumati, ricomposti, e sicuramente non riposti.

Ho letto nella relazione di Patrizia Cotoneschi, che ringrazio per avermi invitata alla conduzione del coordinamento della nostra chiacchierata, che la FUP Florence University Press, è sorta nel febbraio del

¹ Goldschmidt-Clermon, Luisella (1965) "Communication Patterns in High-Energy Physics". In *High Energy Physics Libraries Webzine*(6/2002). <http://library.cern.ch/HEPLW/6/papers/1/>

Il pre-print, che descrive il complesso meccanismo della comunicazione scientifica nei circuiti di pre-preprint per i fisici delle alte energie, è datato febbraio 1965, e per un disguido non venne mai pubblicato rimanendo sepolto nei labirinti della letteratura grigia per trentasette anni, fino alla sua pubblicazione sulla HEPL Webzine del CERN. Per la traduzione italiana dell'articolo, vedi nota 2.

² Goldschmidt-Clermon, Luisella (1965) "Modelli di comunicazione nella fisica delle alte energie" In «Bibliotime», anno VII, numero 1 (marzo 2004). Traduzione italiana dall'originale manoscritto in inglese. per la storia del pre-print vedi De Robbio, Antonella (2004) "Chi ha creato il primo circuito per la distribuzione e lo scambio di preprint?" In «Bibliotime», anno VII, numero 1 (marzo 2004)

2000: potremmo quindi proporre, come ho fatto notare durante l'apertura della tavola rotonda, che la giornata del 10 febbraio 2004 sia in onore del quarto compleanno della FUP.

Ho suddiviso fin da subito, i numerosi interventi della mia tavola rotonda, in quattro sezioni distinte, dove alla fine di ciascuna si è aperto un dibattito "tematico":

- 1) un intervento di apertura, non di overview, ma quasi a carattere normativo, sul deposito legale alla BNCF Biblioteca Nazionale Centrale di Firenze
- 2) due esperienze sul versante dei Data Provider (DP), ovvero l'apertura di depositi, in questo caso istituzionali, in due atenei italiani, Bologna e Trento, per la ricerca e la didattica
- 3) un intermezzo, un aggancio agli argomenti discussi durante la mattinata, focalizzato sul progetto DAFNE, piattaforma per l'editoria elettronica su ambiente OpenSource Zope dell'Università di Padova
- 4) tre interventi orientati all'offerta di servizi a valore aggiunto sui depositi e forniti da terze parti: il primo è un esempio di portale come servizio per l'integrazione degli archivi aperti entro i sistemi bibliotecari, il secondo è basato sulla necessità di creare una piattaforma di servizi a livello nazionale, il terzo è un sistema federato di servizi che agisce ad un metalivello.

Tra un dibattito di fine sessione e l'inizio della sessione successiva ho tentato di tessere le fila di un discorso unitario, fin dove possibile, ma sappiamo che i processi della comunicazione orale che si succedono e si svolgono entro una conferenza non sono mai accadimenti uguali alle successive narrazioni scritte.

Mi viene quasi naturale, vedendo gli appunti disordinati e sparsi tra le mie carte mescolate, ripensare a quanto scritto nel lontano febbraio del 1965 nel pre-print citato sopra, da Luisella Goldschmidt-Clermont, a proposito di preprint e tecniche di comunicazione scientifica, paper riportato alla luce dopo quasi quarant'anni e di cui mi sto occupando in questi giorni.

Alla base della scala comunicativa, troviamo la comunicazione orale, che avviene tra un numero limitato di scienziati, nel nostro caso specifico si trovano presenti, in questa tavola rotonda, alcuni esperti in scienze dell'informazione e biblioteconomia che trattano di cose relative alla scienza. *"Per oscuri motivi"* scriveva Luisella, definita come la prima bibliotecaria di pre-print nel mondo *"una tavola imbandita sembra possedere un certo potere catalitico il cui effetto è anche più grande se vi sono tovaglia e tovaglioli di carta, dal momento che vengono scarabocchiati con tutti i tipi di simboli"*. Questi simboli sono le "mappe concettuali", diremo oggi dopo le teorie di Novak³ sulle tecniche di valutazione dell'apprendimento cognitivo, segni che vanno a costituire ciò che può essere *"determinante nel favorire un alto livello di attività mentale"*.

Non mi soffermerò a descrivere nessuno degli interventi - i quali sono disponibili in questa stessa raccolta di atti - ma inizierò dai miei appunti sparsi e dai fogli scarabocchiati che fortunatamente ho conservato. Attraverso le mie mappe concettuali, disegnate strada facendo durante la tavola rotonda, provo a ricostruire la discussione orale di quel giorno, tentando di trasporla in una forma comunicativa scritta che possa, in qualche modo, concretizzare alcuni percorsi utili ad un lavoro comune per un futuro imminente.

Indubbiamente l'inizio della tavola rotonda pomeridiana subisce il fascino degli echi del dibattito di fine mattina e la tentazione di riprendere da alcuni punti emersi in quella sede è grande, ma è bene iniziare la nostra chiacchierata prima di tutto offrendo alcune nozioni di base.

Che cosa è un archivio e-prints?

E' un Open Archive, o E-prints server, è un archivio preposto al deposito dei documenti scientifici, in forma elettronica, alla loro gestione e conservazione. Può essere usato, come strumento, anche per il deposito dei materiali didattici o delle tesi, anche se nasce come strumento per la disseminazione dei lavori di ricerca. Può essere organizzato a livello istituzionale o a livello disciplinare.

³ Joseph D. Novak della Cornell University. Il termine mappa concettuale è stato coniato da J. Novak e D. Gowin, a partire dalla teoria cognitivista dell'apprendimento significativo, che descrive le strategie per sviluppare e utilizzare questo strumento. La rappresentazione grafica delle conoscenze costringe a riflettere sulla natura delle conoscenze e sulle relazioni che vi intercorrono. Così come una mappa geografica serve per orientarsi in un territorio, una mappa concettuale è strumento per interpretare, rielaborare e trasmettere conoscenze, visualizzando l'oggetto della comunicazione, i concetti principali, i legami che essi stabiliscono e, di conseguenza, il percorso del ragionamento. La teoria nata negli anni sessanta si consolida attorno agli anni novanta

I documenti elettronici possono essere depositati direttamente dagli autori attraverso un semplice processo noto come self-archiving o auto-archiviazione (da non confondere con il self-publishing) o da terze persone (bibliotecari)

Gli e-prints sono copie elettroniche di papers accademici, comprendono: i pre-prints, i post-prints, i post-post-prints. Gli e-prints posti dentro gli archivi di e-prints consentono agli autori di rendere le proprie produzioni di ricerca, liberamente disponibili alla comunità internazionale scientifica, disseminandoli su scala mondiale, cosa impossibile per un lavoro su carta.

Quanto sopra premesso, Giovanni Bergamin della Biblioteca Nazionale Centrale, apre la tavola rotonda sul "*Deposito legale alla BNCF degli archivi di e-prints*", con una citazione di John Kunze sulla persistenza che non è tanto una proprietà dell'oggetto e né della tecnologia, ma deve essere intesa come un problema di servizio, in quanto è impensabile gestire gli archivi senza avere un qualche tipo di informazione sul contenuto ovvero "*impossibile la "persistenza" senza "metadati"*".

Se parliamo di e-prints è naturale che i Service Provider avranno un ruolo fondamentale nell'offerta di servizi di persistenza negli URL, come vedremo nell'ultima sessione della tavola rotonda.

Parlare di deposito ad una tavola rotonda sui depositi ... può anche indurre ad ambiguità interpretative o quantomeno può essere forviante... quindi occorre subito sgombrare il campo da fraintendimenti.

In questa sede si sta discutendo di depositi intesi come archivi o luoghi ove raccogliere (attraverso l'auto-archiviazione) i materiali scientifici e didattici.

Le questioni connesse alla certificazione e tutela vengono spesso mescolate e confuse e, a torto, se ci si riferisce ai depositi di e-prints.

Il deposito legale è una cosa diversa rispetto ad un deposito di e-prints; esso è regolato da una legge che pare sempre sull'orlo dell'approvazione ormai da decenni⁴. Il deposito presso le biblioteche nazionali, stabilito per legge, di una copia di quanto pubblicato garantisce un accesso all'informazione, ma stiamo parlando pur sempre di "pubblicazioni".

Bergamin esordisce citando l'editto del 1537 di Francesco I che sanciva che "*nessun libro può essere messo in circolazione senza un esemplare depositato nella biblioteca del castello reale di Blois*"

Entro i "depositi" istituzionali, va ricordato, convive anche materiale cosiddetto "grigio", pre-print, report tecnici, report interni, working papers, ecc...

E' opportuno che una copia di detto materiale "non ancora pubblicato" e che in certi casi non sarà mai pubblicato, sia conservato in un deposito nazionale per legge?

In qualche modo questo "deposito legale" può favorire una qualche forma di "certificazione" del documento pre-print o comunque sicuramente del post-print? Sappiamo bene come la questione "formale" della certificazione sia uno dei punti spinosi nello stabilire il valore delle pubblicazioni elettroniche nei concorsi universitari.

Il ruolo delle biblioteche nazionali non cambia con l'avvento delle pubblicazioni elettroniche.

E dal pubblico giungono altre domande⁵: "chi" seleziona "cosa" conservare? E "chi" conserverà "cosa"? Quali soggetti saranno coinvolti? Verrà conservato tutto ciò che si pubblica, o solo una parte della letteratura permanente? Per quali tipologie di documenti è previsto il deposito?

Ma come in tutti i convegni che si rispettino, non c'è mai tempo per sviscerare a fondo le questioni, perché altre si vanno via via affastellando ...

Quello che potremmo fare è stabilire alcune premesse fondamentali, sorte a seguito anche di alcune domande provenienti dal pubblico⁶, durante la mattinata, che avevano generato una qualche incertezza o perlomeno alcune perplessità e sulle quali ora posso concedermi di puntualizzare in modo meno sommario rispetto alle mie premesse orali:

- differenziazione della questione **certificazione**, dove potremmo individuare, dalla nostra mappa concettuale, tre differenti aspetti della certificazione:

⁴ Legge 2 febbraio 1939, n. 374 ("Norme per la consegna obbligatoria di esemplari degli stampati e delle pubblicazioni") pubblicata nella Gazzetta Ufficiale 6 marzo 1939, n. 54 e modificata in piccola parte dal decreto luogotenenziale n. 660 del 31 agosto 1945.

⁵ Ringrazio il Prof. Riccardo Ridi dell'Università di Venezia, che dal pubblico è intervenuto alla nostra tavola rotonda dando spunti notevoli per il dibattito; le sue osservazioni a seguito dell'intervento di Bergamin mi sono state preziose per l'analisi di seguito esposta sul tema del deposito legale

⁶ Mi riferisco in particolare agli interventi del Prof. Antonio Fantoni coordinatore del CITICORD (Centro Interdipartimentale per le Tecnologie dell'Informazione e della Comunicazione nella Ricerca e nella Didattica) dell'Università Roma La Sapienza che ha toccato gli aspetti della certificazione, qualità e tutela, aspetti tutti molto rilevanti nel suo settore disciplinare, la genetica

- pertinente al documento in sé, come atto riconosciuto ufficialmente, e questo aspetto andrebbe definito più correttamente come **autenticazione** (quale versione del papers sto effettivamente leggendo?). In questo contesto rientrano anche tutti gli aspetti relativi all'integrità e fedeltà del documento, ovvero la garanzia che non sia un documento corrotto o in qualche modo contraffatto.
- correlato all'esigenza di **certificazione di qualità** del contenuto ai fini della carriera (forma e contenuto sono anche qui due aspetti distinti). Questo aspetto riguarda le attività di peer-review e rientra nel canale dei periodici con comitato editoriale.
- necessità di stabilire la priorità della scoperta scientifica: in tale aspetto la certificazione come momento zero serve a stabilire la **priorità intellettuale** ed è per questo anche anello di congiunzione con la sfera della tutela.
- separatezza della questione relativa alla certificazione di cui al punto precedente, rispetto a quella della **tutela** che ricade in altra sfera, ossia nella sfera della **proprietà intellettuale** e qui il discorso potrebbe divenire complesso. Inizio e mi fermo con l'asserzione che la tutela è finalizzata alla protezione dell'opera e che la tutela è un aspetto normativo e non tecnico, né tantomeno organizzativo e che comunque la disseminazione informativa viaggia in canali indipendenti. In questa tavola rotonda la questione del copyright non viene trattato da nessun relatore, anche se siamo tutti consapevoli che uno degli ostacoli alla crescita ed espansione degli archivi di e-prints è proprio la questione copyright.

Il deposito legale si colloca nella prima sfera e comunque non entra nel merito della certificazione di qualità dei contenuti, ma si "ferma", nel caso del materiale e-prints all'"*adempimento meramente burocratico per radicare la sussistenza del titolo scientifico*" come ben ci dice Brugaletta, quando afferma che:

*1. la disciplina prevista dalla legge e richiamata nei bandi e la stessa giurisprudenza [...] si può riferire solo alla pubblicazioni a stampa e non alle pubblicazioni effettuate con altri mezzi o supporti (la stessa giurisprudenza che esclude dal campo delle pubblicazioni gli interventi e le comunicazioni ai congressi non può costituire un precedente negativo nei confronti di forme di pubblicazioni, come sono quelle elettroniche e quelle nel web, che hanno identiche finalità di quelle su carta e un campo di azione addirittura superiore); 2. le pubblicazioni effettuate con mezzi diversi dalla stampa sono suscettibili di essere valutate senza la necessita di osservare le formalità previste per i lavori a stampa;*⁷

Certamente forme di deposito legale per gli e-prints entro un modello distribuito, previste sin da ora, potrebbero venire in soccorso alla riformulazione di una nuova disciplina in materia, considerando anche che gli stessi depositi istituzionali potrebbero giocare un ruolo strategico nello scenario laddove il deposito di lavori relativi a scoperte scientifiche "certificherebbero" le priorità intellettuali delle ricerche svolte negli atenei e negli enti di ricerca.

La garanzia inoltre di "persistenza" rafforzerebbe il cammino verso la certificazione.

Proseguendo nella nostra tavola rotonda, Serena Spinelli illustra l'esperienza degli "Archivi e-prints dell'Università di Bologna".

Paolo Bellini e Francesca Valentini dell'Università di Trento, descrivono, subito dopo, le fasi del progetto "Unitn-Eprints: risultati di utilizzo, prospettive di sviluppo".

Il software usato da entrambi gli atenei è Eprints, lo stesso usato anche all'Università di Firenze e, sebbene ancora in fase sperimentale, all'Università di Padova.

Non mi dilungo sulle caratteristiche di amichevolezza e flessibilità di questo software OpenSource perché ne parleranno i relatori, ma voglio sottolineare che tale software, nato nel febbraio di due anni fa, è adottato da 123 archivi nel mondo contenenti ad oggi oltre 39.000 documenti.

Pur nella loro diversità di approccio le due esperienze denotano percorsi comuni e gli stessi sono i problemi incontrati nel trovare un consenso generalizzato al fine di riempire questi depositi istituzionali.

Ciò che emerge chiaramente è che non sono tanto le scelte tecnologiche a strutturare il servizio, ma le scelte organizzative in merito ad alcuni punti fondamentali:

- quanti depositi aprire entro un sistema

⁷ Brugaletta, Francesco, "Brevi note sul valore delle pubblicazioni elettroniche nei concorsi". 08.01.98
<http://www.interlex.it/accesso/brugal9.htm>

- quale tipologia di documenti inserire (cosa si intende gestire e trattare: tesi, paper di ricerca, materiale didattico, conferenze ...)
- quale struttura gerarchica utilizzare come nucleo per la generazione di “viste” per il browsing dell’archivio (settori disciplinari ex MURST, lista gerarchica per Facoltà, schemi di classificazione scientifici o biblioteconomici, ...)
- quali politiche di sottomissione all’archivio adottare: chi deposita cosa e in quali formati e come si deposita (in modo autonomo attraverso il self-archiving o attraverso uno staff o un servizio di help-desk)
- le politiche sul copyright: relazioni tra autore/editore (cessione di diritti), relazione tra istituzione/autore
- le questioni di layout per la presentazione “ufficiale” del sito, compreso loghetto e nome del server
- in quale lingua si presenta il server (solo italiano, inglese, entrambi? in tal caso vanno previsti gli allineamenti ad ogni nuova versione)
- le configurazioni per l’accesso ai servizi: profili di utenze, servizi a valore aggiunto in connessione con i Service Provider, ricerca a testo pieno
- la definizione di metadati per ciascuna tipologia di materiale previsto (con apposite mappature utili all’harvesting)
- la validazione del server entro l’architettura OAI ai fini della visibilità internazionale
- come trovare consenso alla base (promozione allargata dal basso) e come trovare consenso ai vertici entro gli organi ufficiali accademici (riconoscimento dell’archivio come strumento ufficiale)

Ma se è difficile trovare consensi tra il corpo docente, non è nemmeno facile far comprendere ai colleghi bibliotecari che gli e-prints fanno ormai parte del nostro processo evolutivo. Ciò che si vede sono solo i “difetti” di questi strumenti ancora neonati, senza comprendere che bisogna dar loro il tempo per crescere; nella ricerca dei dettagli e delle reali o presunte lacune dello strumento si perdono di vista i numerosi vantaggi che deriverebbero da un massiccio utilizzo dei depositi. Tale miopia, soprattutto da parte dei bibliotecari, ci porta a consolidare meccanismi controproducenti che ci succhiano ogni anno ingenti risorse destinate alla ricerca. Invece che aiutare gli archivi di e-prints a crescere si puntano tutti gli sforzi nel negoziato con gli editori in consorzi per l’acquisto di periodici elettronici⁸ stabilizzando il perverso meccanismo tradizionale basato sulla sottoscrizione di materiale che ci appartiene in quanto da noi prodotto.

Qualcuno del pubblico poneva la giusta osservazione sugli strumenti di authority, tesauri e schemi di classificazione, attualmente assenti nello scenario degli archivi di e-prints.

Va detto a proposito che un archivio per la ricerca e la didattica non è un database catalografico, ha scopi e obiettivi diversi rispetto ad un catalogo, in quanto si fonda e nasce su presupposti diversi.

E non è nemmeno un catalogo nel senso più profondo del termine, non è un simil-MARC per intenderci; i suoi metadati sono di basso livello, nella maggior parte degli archivi si usa Dublin Core non qualificato (s-qualificato si potrebbe scherzosamente definire) o al massimo Dublin Core con qualificatori come nel caso delle esperienze con la piattaforma DSpace.

Per tale ragione strumenti per l’authority file, tesauri, o schemi di classificazione sono demandati ai Service Provider i quali potremmo definirli come servizi a valore aggiunto costruiti attorno, dietro, davanti, in mezzo, sopra, i Data Provider... ad un livello meta di aggregazione.

Un archivio di e-print per la ricerca nasce con lo scopo di registrare le produzioni intellettuali dei propri docenti e ricercatori e rientra in uno dei due canali dell’Open Access. L’Open Access è una filosofia e ha lo scopo di arginare l’emorragia di spesa per i periodici in generale (cartacei ed elettronici) che toglie fondi alla ricerca. Il suo obiettivo primario è quello di riappropriarsi delle produzioni intellettuali di ricerca esternalizzate nel corso degli anni agli editori/oligopoli.

Un archivio e-prints può servire come strumento per la valutazione della ricerca dell’ente, nel nostro caso l’ateneo, non è uno strumento che fa qualità, per questo ci sono i periodici e mi ricollego alla questione della certificazione. Potrebbe per esempio essere visto come strumento per la certificazione utile a stabilire la proprietà intellettuale, tramite il campo interno etichettato come “DataStamp”.

Ma cosa offrono in più i depositi istituzionali rispetto ai depositi disciplinari centralizzati?

⁸ Numerose biblioteche nel mondo tra cui il servizio delle biblioteche di Cornell (Stati Uniti) invece ha scelto di eliminare parecchi abbonamenti alle riviste scientifiche del gruppo Elsevier, spiegandone le ragioni in una lettera aperta.

<http://www.library.cornell.edu/scholarlycomm/elsevier.html>

- offrono una risposta strategica ai problemi esistenti nel sistema che regola i periodici scientifici (dai costi ormai insostenibili)
- offrono un immediato complemento di qualità al modello di pubblicazione scientifica ad oggi esistente
- stimolano l'innovazione entro una struttura di editoria disaggregata
- servono come indicatori tangibili della qualità di un'istituzione, ne incrementano la visibilità, il prestigio e il valore a livello pubblico

Un'altra premessa di sfondo è quella relativa alla confusione che esiste oggi tra depositi e-prints ed editoria elettronica, e ancora peggio tra editoria elettronica open access (sostenibile) e university press. Si tratta di segnare i giusti confini perché possono variare i modelli economici sviluppati sulla base di scopi e mission ben precisi.

Nel canale Open Access convivono, come abbiamo detto, due strategie complementari

- *Open-Access Journals* dall'altra: è il canale ove sono collocati i periodici che non fanno pagare abbonamento o tasse per l'accesso online, ma che prevedono altre fonti di finanziamento per il referaggio o la pubblicazione
- *Self-Archiving* da una parte: gli scienziati dovrebbero essere in grado di depositare i propri articoli "refereed journal articles" in archivi aperti conformi allo standard Open Archives Initiative

Sono quindi due i canali dell'Open Access:

- Pubblicazioni entro iniziative di *editoria elettronica sostenibile, o comunque Open Access*
- Deposito negli Open Archives

Noi nella nostra tavola rotonda ci occuperemo solo di questo secondo canale, sebbene torni utile ricordare come le due strade si incrocino. A livello funzionale i due canali sono strettamente correlati e qui ci riconnettiamo alla questione della certificazione.

Le funzioni chiave di un periodico, come descritto dal mitico lavoro di Roosendaal e Geurts del 1997⁹ sono quattro. registrazione, certificazione, consapevolezza, archiviazione e vengono assicurate in modo differenziato

- dagli **archivi istituzionali** per
 - la **registrazione** che offre un valido supporto per stabilire la priorità intellettuale di una scoperta e, in un certo senso, agendo come strumento ai fini della tutela sul contenuto
 - la **consapevolezza** che include l'accessibilità e la disseminazione, fondamentale momento di scambio per la rielaborazione di idee e concetti condivisi, ai fini della produzione di nuova conoscenza
 - l'**archiviazione**, intesa a due livelli: come semplice anagrafe della produzione intellettuale dell'ente o, nei server più evoluti, come archivio preposto alla conservazione a lunga memoria
- dalle **periodici open access** per
 - la **certificazione**, che viene assicurata attraverso i meccanismi di peer-reviews consolidati, da gruppi di esperti (i pari) che validano i papers

<p>Registrazione: stabilire la priorità intellettuale di un'idea, un concetto, o una scoperta scientifica</p> <p>Archivi istituzionali</p>	<p>Certificazione: certificare la qualità della ricerca e/o la validità delle scoperte dichiarate the;</p> <p>Periodici open access</p>	<p>Consapevolezza: assicurare la disseminazione e l'accessibilità delle produzioni della ricerca, fornendo un mezzo attraverso il quale i ricercatori possono essere avvertiti delle novità nel campo</p> <p>Archivi istituzionali</p>	<p>Archiviazione: conservare il patrimonio intellettuale per le fruizioni future</p> <p>Archivi istituzionali</p>
--	---	--	---

⁹ Roosendaal, Hans E. and Peter A.Th.M. Geurts (1997). Forces and functions in scientific communication: an analysis of their interplay. available at <http://www.physik.uni-oldenburg.de/conferences/crisp97/roosendaal.html>.

Potremmo forse anche parlare di metamorfosi a seguito di un certo numero di editori tradizionali che stanno trasformando i loro periodici ad accesso chiuso in open access journals, tra i quali spiccano nomi prestigiosi come:

- Oxford University Press
- Company of Biologists
- American Physiological Society
- Florida Entomological Society
- Entomological Society of America
- Infotrieve

Il "Progetto DAFNE: integrazione degli archivi e-prints nel circuito dell'editoria elettronica italiana" come momento di concreta realizzazione di una piattaforma per l'editoria elettronica su ambiente Open Source Zope., ci viene descritto da Ornella Volpato del Centro di Ateneo per le Biblioteche (CAB) dell'Università di Padova.

A livello tecnico la piattaforma Zope per DAFNE¹⁰ è simile ad un deposito, ma non dobbiamo confondere le potenzialità insite in questo strumento che consente di gestire la complessa rete di relazioni tra referi e membri dei comitati editoriali, tra autori e comitato stesso, tra comitato editoriale e le altre figure che intervengono nel complesso panorama della costruzione di un periodico a visibilità internazionale.

Finanziato dal MURST allo scopo di offrire assistenza a università e associazioni accademiche per lo sviluppo e disseminazione della comunicazione scientifica, DAFNE è un progetto mirato alla valorizzare dell'editoria scientifico - accademica italiana e si propone di favorirne la trasformazione in editoria elettronica.

L'obiettivo principale di DAFNE è realizzare un prototipo di infrastruttura nazionale per le pubblicazioni elettroniche (principalmente periodici). Un prototipo del sistema previsto è stato sviluppato utilizzando come campione diversi documenti nell'area delle scienze sociali. Si tratta di una infrastruttura ad ampio spettro in cui i metadati per la descrizione e gestione delle risorse e dei processi in cui esse sono coinvolte hanno un ruolo essenziale.

Allo scopo sono presi in considerazione tutti gli aspetti tecnologie organizzativi per una disseminazione delle pubblicazioni italiane (principalmente nei campi delle scienze umane) a livello internazionale.

Dafne intende realizzare un prototipo nazionale di infrastruttura tecnologica e organizzativa per i prodotti dell'editoria elettronica, ovvero: un "distretto industriale" che immetta in un circuito complessivo la domanda e l'offerta culturale digitale. Il modello tecnico e architettonico si basa sullo standard OAI *Open Archival Information System* e prevede l'interazione fra i fornitori di dati, i fornitori dei servizi e vari servizi esterni.

DAFNE individua il settore chiave della produzione scientifico-accademica nei periodici e nel loro corrispettivo periodico elettronico. Quest'ultimo supporto garantisce la maggior dinamicità del mercato editoriale alla condizione che vengano rispettate in via preliminare alcune condizioni di formato, agibilità e collegabilità ad altri oggetti culturali in una struttura ipertestuale facilmente utilizzabile e condivisibile dall'utente finale. A partire dai due modelli economici oggi esistenti per l'accesso all'informazione digitale, che vedono protagonisti gli editori commerciali e quella parte del mondo accademico aperto alla libera diffusione delle pubblicazioni accademiche, il progetto propone servizi in cui ambedue le modalità siano rese possibili, prevedendo varie funzionalità.

Con la piattaforma su Zope si vuole in sostanza proporre uno strumento che possa:

1. Assistere il comitato editoriale nei processi di refereeing e abstracting attraverso un'agevole comunicazione tra le varie figure del circuito
2. Assistere gli autori nella pubblicazione elettronica dei loro papers offrendo un servizio editoriale di qualità
3. Valorizzare la produzione editoriale dell'Ateneo nel rispetto dei diritti di proprietà intellettuale.
4. Proporre un modello di riferimento di editoria elettronica accademica utile ad essere applicato ad altre pubblicazioni del nostro Ateneo

¹⁰ District Architecture for networked edition: technical model and metadata.

<http://www.bncf.net/dc2002/program/ft/paper1.pdf>

In merito a cooperazione e integrazione DAFNE è la sintesi di tre complessi progetti studiati rispettivamente dalle Università di Bologna, Firenze e Padova tesi a realizzare alcuni significativi segmenti di un articolato prodotto / processo riconducibile alla cosiddetta "biblioteca digitale".

Allo scopo si sta effettuando un test-bed con la rivista Rendiconti del Seminario Matematico dell'Università di Padova.

Liliana Morotti di ATLANTIS s.r.l, tratta "L'integrazione degli e-prints fra le risorse di Metalib". L'integrazione degli e-prints entro il sistema informativo di ateneo e il conseguente recupero dell'informazione dal portale di ateneo è una questione strategica ai fini di un impatto positivo sull'utenza in termini di immagine. Per tale ragione sarà strategico a mio avviso definire alcune linee di intervento che prevedano:

- la perfetta integrazione dei lavori a testo pieno contenuti negli e-prints con le corrispettive versioni pubblicate sul sito degli editori, attraverso strumenti di reference-linking come Sfx
- l'inserimento nei cataloghi di tutti i periodici open access per un recupero "prioritario" degli articoli contenuti rispetto agli articoli pubblicati su riviste chiuse a pagamento: intendo dire che in fase di ricerca via MetaLib dovrebbero comparire per primi i contenuti open access rispetto a quelli toll access

Questo perché in certi ambiti come la Fisica per il 72% degli articoli pubblicati in *Astrophysical Journal* è disponibile almeno una versione free access (principalmente su ArXiv). Inoltre questo 72% di articoli, in media, si cita due volte più spesso del restante 28% che non ha una versione libera corrispettiva. Potrebbe anche accadere e ce lo auguriamo tutti che, nei prossimi anni, si assista ad una radicale trasformazione dei meccanismi di valutazione della ricerca che ora, come ben sappiamo, si basano sull'indicatore bibliometrico noto come IF Impact Factor, non sempre adeguato a rispondere agli scopi e ai bisogni di tutte le comunità scientifiche.

Una maggiore attenzione ai downloads degli articoli (e dei lavori sugli e-prints naturalmente), attraverso attento monitoraggio di log, accessi e scarico di file, potrebbe condurre a risultati sorprendenti. Peter Suber, nel blog di Open Access News¹¹, asserisce che la media del numero dei downloads per articolo in ScienceDirect negli ultimi anni è stata di **28**. Nello stesso periodo la media di downloads per gli articoli in BioMedCentral è stata di **2.500**, mentre per il *New Journal of Physics* la media è stabilizzata da alcuni anni a quota **1000**.

L' Open Access, offre un utilizzo pari a **89 volte per la biomedicina e a 35 volte per la fisica** rispetto al modello chiuso tradizionale basato sulla sottoscrizione.

Fin qui abbiamo parlato di ricerca, ma daremo uno sguardo veloce anche all'ambito della didattica.

Considerato che il nostro soggetto sono gli archivi, un'altra questione emersa durante la mattinata¹² e su cui è bene puntualizzare a scanso di equivoci, concerne il rapporto tra archiviazione dei materiali didattici e piattaforme e-learning e conseguente utilizzo di metadati. Va detto che attualmente le piattaforme per la didattica online, o piattaforme e-learning sono completamente sganciate dai depositi istituzionali, siano esse costruite con software commerciale, o basate su software libero o a codice sorgente aperto.

A livello internazionale - considerato che si tratta di una situazione comune e purtroppo generalizzata - servirebbe agire verso l'integrazione di questi ambienti - ora territori a completo dominio della docenza - riconducendoli entro i sistemi bibliotecari, soprattutto, ma non solo, ai fini di un'offerta di servizi bibliotecari a supporto della formazione a distanza. Non dobbiamo dimenticare il recente decreto sulle università telematiche¹³. Le piattaforme e-learning andrebbero dotate di metadati standard per la descrizione del materiale depositato sia dai docenti, sia dagli utenti (nelle aree di apprendimento collaborativo). Va anche ricordato che nei corsi e programmi online, sempre più spesso, si creano e si utilizzano materiali didattici prima sconosciuti, i Learning Object (LO) i quali necessitano di opportuni metadati Learning Object Metadata (LOM) per essere non solo descritti (metadati descrittivi), e gestiti (MAG), ma anche nella fase di progettazione creazione e di riutilizzo dell'oggetto di apprendimento. In ogni caso tali ambienti, dovrebbero essere integrati non solo con gli altri database catalografici o di altra natura, ma dovrebbero essere in sinergia con depositi di materiali didattici, compresi i cosiddetti Learning Object Repositories (LOR), particolari tipologie di depositi contenenti oggetti per l'apprendimento finiti o contenenti il materiale grezzo precedente lo "storyboard" dell'oggetto stesso (in questo caso depositi utili al riuso dei LO).

¹¹ Precedentemente FOS Free Online Scholarship <http://www.earlham.edu/~peters/fos/fosblog.html>

¹² Sollevata da Paolo Gardois dell'Università di Torino

¹³ Decreto sulle Università telematiche del 17 aprile 2003. Criteri e procedure di accreditamento dei corsi di studio a distanza delle università statali e non statali e delle istituzioni universitarie abilitate a rilasciare titoli accademici. (Gazzetta Ufficiale N. 98 del 29 Aprile 2003)

Chiudo questo inciso sui LO, LOM e LOR e sugli ambienti e-learning perché questo richiede una sede di discussione diversa, sebbene spesso accada che nelle sedi opportune mai si discuta (per ovvie non-ragioni) dell'integrazione tra depositi/archivi e piattaforme, di metadati in integrazione con gli altri metadati entro il sistema informativo di ateneo e qui i portali potrebbero giocare un ruolo determinante.

Ecco ...siamo ormai entrati nella quarta e ultima sessione della nostra tavola rotonda, la più complessa per la dinamicità degli interventi, come abbiamo visto dall'intervento precedente sull'integrazione entro i portali. Mi preme sottolineare ancora una volta, anche se forse può sembrare scontato, che l'architettura dell'Open Archive Initiative è suddivisa in due parti: da una parte ci sono i Data Provider, dall'altra i Service Provider. Nei DP abbiamo dati, contenuti a testo pieno legati ai propri metadati raccolti e indicizzati dai SP. E' un'architettura agile e basata su un modello distribuito che ben si presta all'implementazione di strumenti a valore aggiunto quali appunto schemi di classificazione e liste di autorità, anche se in tale direzione molto si dovrà lavorare in cooperazione con chi produce dati e metadati (i DP).

Ciò che serve a livello di cooperazione italiana è una piattaforma per l'accesso agli strumenti software OAI, che sia di supporto alla costruzione di archivi per la ricerca, ma anche per la didattica, istituzionali o disciplinari, alla creazione di servizi per la raccolta e il caching di metadati, harvester, per il reference linking entro gli Open Archives, per conversione di metadati da un formato all'altro, per l'esportazione e importazione di metadati, per la ricerca nel testo pieno dei documenti. Tutto questo rientra nella sfera di competenza dei Service Provider.

Susanna Mornati del Cilea ha presentato il "Progetto AEPIC: gli Archivi aperti italiani su una piattaforma nazionale" che va in questa direzione.

Il progetto *COAP CILEA Open Archives Platfor*, sottoprogetto di AEPIC, è basato su progetti simili di altri Paesi, come il britannico UK e-prints¹⁴ e l'australiano. E' un'iniziativa che nasce come volontà di collaborazione con le università e istituzioni scientifiche e come momento di aggregazione delle varie comunità di ricerca italiane al fine di poter offrire un accesso esteso e collettivo alle produzioni intellettuali della ricerca in Italia. *COAP* si occupa della creazione di un'infrastruttura italiana di repositories per la ricerca entro un contesto di editoria digitale Open Access offrendo supporto, strumenti e spazi, a quanti (persone o istituzioni) sono interessati a soluzioni open access per la disseminazione dei contenuti di ricerca. *COAP* vuole inoltre essere principalmente un servizio basato sulla raccolta di metadati dagli archivi istituzionali o disciplinari italiani, entro un database centralizzato che consenta di raggiungere agevolmente il testo pieno del lavoro di ricerca depositato in uno dei server distribuiti nella rete. Al centro della piattaforma trova spazio il Service Provider con l'harvester per la raccolta dei metadati dai vari archivi ubicati dentro e fuori la piattaforma unitamente all'archivio di metadati centralizzato dove viene creato l'indice.

Attorno, quasi a formare una costellazione, gli archivi disciplinari che trovano spazio entro *COAP*. A modello esiste E-LIS¹⁵, progetto pilota, già operativo e funzionante, un Open Archive internazionale costruito con software EPrints, organizzato a livello disciplinare per l'ambito LIS Library and Information Science, già dotato di numerose funzionalità e applicazioni OAI, comprensivo ad oggi di quasi 700 documenti a steso pieno.

La piattaforma è adatta anche ad ospitare server istituzionali per materiale di ricerca (software EPrints) o per materiale misto, ricerca e didattica (software DSpace, come nel caso del server dell'Università di Parma).

Pasquale Pagano del CNR di Pisa ci parla dello sviluppo di "OpenDLib: un sistema federato di servizi di biblioteca digitale su archivi" da parte del Gruppo DLib, guidato da Costantino Thanos del ISTI Institute del Consiglio nazionale della Ricerca – Italia.

OpenDLib è uno strumento software che può essere usato per generare facilmente una biblioteca digitale, secondo i requisiti di date comunità di utenti, fornendo appropriatamente i dati al software e alternativamente caricando o raccogliendo i contenuti da gestire.

OpenDLib, appare come una federazione di servizi, ognuno gestito da una specifico "server". L'intera architettura può essere replicata, oppure distribuita in diversi punti della rete. I vari "server" comunicano fra loro attraverso un protocollo basato su http, il sistema può quindi essere usato anche in contesti dove tutti i metadati, o parte di essi, sono memorizzati in un qualsiasi "Repository" a condizione che questo sia provvisto di una interfaccia OAI.

¹⁴ <http://www.rdn.ac.uk/projects/eprints-uk/>

¹⁵ <http://eprints.rclis.org>

Il progetto mira a strutturare una infrastruttura per digital library al fine di supportare la comunicazione e la collaborazione tra le reti delle varie comunità scientifiche. Inoltre mira alla fornitura di standard per acquisizione, descrizione, archivio, accesso, ricerca e disseminazione dell'informazione. Questa infrastruttura può supportare dati di tipo non testuale, annotazioni, ricerche cross-language e disseminazione personalizzata dell'informazione

In sostanza OpenDLib consiste di un sistema di servizi che implementa le funzionalità di una biblioteca digitale compiendo poche assunzioni sulla natura dei documenti da immagazzinare e da diffondere. Se necessario, il sistema può essere esteso con altri servizi per andare incontro a particolari esigenze.

OpenDLib è stata sviluppata come una biblioteca digitale distribuita, in accordo con la nozione di servizi individualmente definiti dislocati ovunque in Internet. Quando combinati, questi servizi costituiscono una biblioteca digitale.

Queste due ultimi interventi focalizzati su esperienze sul versante dei servizi di tipo SP OAI sono strettamente integrati. Per esserci impatto è necessaria un'ampia disseminazione. I lavori dei ricercatori devono essere letti, citati e utilizzati da altri ricercatori: solo così raggiungono l'impatto utile ad uno sviluppo collettivo. Per questa ragione occorre fare attenzione ai contratti editoriali nel trasferimento dei diritti (copyright) agli editori: ritenzione di alcuni diritti, cessione non esclusiva

Sarà quindi necessario mettere in piedi un gruppo di lavoro italiano che lavori su queste tematiche basandosi prevalentemente sui documenti prodotti dal progetto europeo RoMEO (*Rights on METadata for Open archiving*)¹⁶, che era sorto per investigare la questione dei diritti che ruotano attorno l'auto-archiviazione per i lavori di ricerca depositati negli Open Archive della comunità accademica britannica. Il progetto tiene conto delle questioni correlate alla raccolta dei metadati dai data providers entro i service providers entro il quadro OAI. Sono stati sviluppati una serie di elementi sui diritti a partire dalla costruzione da esistenti schemi e vocabolari (Open Digital Rights Language) da inserire entro i metadati. Obiettivo di Romeo è stato quello di creare un sistema che dimostri come i diritti sui metadati possono essere assegnati, scoperti, raccolti e visualizzati agli utenti attraverso il Protocollo OAI per la raccolta dei metadati. Le indagini condotte sugli autori nell'ambito del progetto hanno dato risultati assai interessanti.

In sostanza due sono le aree in cui i soggetti coinvolti dovranno imparare a collaborare.

Da una parte abbiamo il ruolo degli atenei e delle istituzioni con i data provider o archivi di contenuto.

Dall'altra abbiamo il ruolo di terze parti come i consorzi, gli editori, i commerciali che dovranno impostare le loro offerte di servizio per il potenziamento dell'infrastruttura di ricerca.

Obiettivo comune e condiviso è quello di favorire la crescita di depositi istituzionali per le produzioni intellettuali della ricerca in Italia disponendo di una rete di servizi a supporto dei soggetti (istituzionali e non) interessati all'apertura e costruzione di archivi aperti per la ricerca.

Altro obiettivo parallelo è quello di cooperare tutti assieme ad incrementare il numero dei lavori di ricerca depositati dentro gli archivi aperti cercando di rimuovere le barriere che attualmente ostacolano il processo di self-archiving.

Le campagne di PLoS hanno mosso alcuni interessi politici, vedi negli Stati Uniti il recente atto normativo noto come Sabo 'Public Access to Science' Act, che prende il nome dal suo promotore Martin Olav Sabo¹⁷ e che stabilisce che i dati e le ricerche finanziate con fondi pubblici devono essere accessibili in quanto '*...an optimum international exchange of data, information and knowledge contributes decisively to the advancement of scientific research and innovation*' and '*...open access will maximise the value derived from public investment in data collection efforts.*'

Dovremo anche in Italia muoverci a questi livelli per intervenire nella scrittura di leggi e decreti e qui il ruolo della CRUI dovrebbe essere pro-attivo piuttosto che passivo.

¹⁶ Informazioni sul Progetto Romeo (Rights METadata for Open archiving)

<http://www.lboro.ac.uk/departments/ls/disresearch/romeo/index.html>

Romeo è stato coordinato da Charles Oppenheim dell'Università di Loughborough, e si è occupato di indagare come i metadati e i dati (lavori originali di ricerca) 'give-away research literature' sono utilizzati e come questi possono essere protetti. Il progetto ha terminato i suoi lavori, ma continuerà a svilupparsi in coordinamento con l'OAI.

◆¹⁷ <http://www.biomedcentral.com/news/20030627/04>

Il valore aggiunto sta nei seguenti servizi:

- Motori di ricerca per le ricerche nei full-text: vedi esperienze CERN e Caltech
- Analisi delle transazioni Log e monitoraggio dei downloads: statistiche sugli accessi al server, ma anche sul singolo autore e su ogni singolo articolo (vedi esempio ArXive, RePEc per abstract visualizzati, articoli scaricati, per giorno, mese, anno ...)
- Persistenza degli URL
- Analisi scientometriche e webmetriche: nuovi strumenti per il calcolo e le nuove misurazioni dell'impatto di ricerca, per il recupero e linking delle citazioni nel web (vedi Paracite, OpCite, CiteBase, Citeseer/ResearchIndex)
- Help-desk sulle questioni correlate al copyright e adozioni di politiche comuni sui diritti: supporto agli autori per il contratto editoriale. Attenzione alla cessione dei diritti! Perché regalare la proprietà intellettuale all'editore? Sul contratto editoriale si può intervenire. Non occorre essere sempre passivi nei confronti degli editori. Andare a vedere contratti "favorevoli" sul sito di RoMEO.
- Zona "strumenti" di tipo opensource, liberamente accessibile dalla comunità OAI italiana

Siamo alla fine della nostra giornata... mancano alcuni attori importanti qui oggi seduti alla nostra tavola rotonda, anche se sono presenti tra il pubblico e mi riferisco alle esperienze dell'Università di Messina in fase di decollo, a Roma La Sapienza in fase di progettazione, al CNR di Bologna che ha sviluppato SAIL, un Service Provider con harvester di tutto rispetto e probabilmente molte altre ...

Il valore aggiunto dei servizi risiede soprattutto nella collaborazione tra questi differenti soggetti, in una distribuzione di compiti secondo competenze chiare e definite, in un lavoro che finalmente riesca a coniugare le professionalità di ciascun soggetto a favore di progetti nazionali che diano al nostro Paese la visibilità che merita e che oggi non ha a causa degli sforzi notevoli ma frammentati e non coordinati.

Una lista di discussione Web sul sito del progetto CILEA, appositamente dedicata sia agli aspetti tecnici sia agli aspetti sociologici e culturali, è già operativa.

Anche se in chiusura è doveroso però soffermarci un attimo sui rischi che i nuovi modelli economici in tema di editoria elettronica accademica potrebbero comportare. Ciò è un po' fuori obiettivo rispetto al tema della nostra tavola rotonda che si è occupata dei depositi di e-prints.

Poiché i depositi e-prints rientrano in uno dei due canali dell'Open Access e poiché oggi abbiamo anche sfiorato il secondo canale, quello dell'editoria open access sostenibile, mi sembra corretto citare l'aspra polemica che si sta conducendo in questi giorni entro le liste internazionali tra cui American Scientists.

Will 'Open access' be open to everyone? E' la domanda che corre frenetica tra le liste di discussione di tutto il mondo a seguito della lettera di John Ewing¹⁸, matematico dell'AMS, che ha recentemente scatenato una serie di polemiche sul modello Open Access considerato da alcuni studiosi poco "testato" e rischioso ...Ewing nella sua lettera argomenta che i periodici open access escludono gli autori poveri che non possono permettersi di pagare la loro quota per la pubblicazione. Ewing punta in particolare al piano PLoS il qual prevede una tassa per gli autori che vogliono pubblicare e conclude dicendo che ciascun modello di pubblicazione sia esso basato sulla sottoscrizione (attuale) sia sostenuto dall'autore o dall'istituzione (open access) ha i suoi vantaggi e svantaggi che però non sono perfettamente bilanciati. Quando uno scienziato non è abbonato ad un periodico, può in ogni caso ottenere l'informazione sull'articolo che cerca o da una lista di riferimenti bibliografici o perché trova l'abstract in linea. Comunque può sempre richiedere una copia dell'articolo, anche via mail, ad un collega o tramite il servizio di document delivery fornito dalle biblioteche o, alla peggio, rivolgendosi ad un centro di fornitura documenti a pagamento. Quando invece uno scienziato non ha i finanziamenti per pubblicare un articolo, quell'articolo non comparirà mai, non solo come riferimento bibliografico, ma, peggio ancora, non verrà mai pubblicato e non farà mai parte della letteratura permanente.

Questo, dice Ewing, è molto di più che un inconveniente...

Vanno perciò individuati modelli economici adatti per nuove forme di editoria elettronica accademica che affianchino i depositi di e-prints in modo efficace e che prevedano un'attenta analisi su costi e benefici, senza incorrere in rischi che comportino inaccessibilità ben più pesanti delle attuali già gravemente compromesse dagli oligopoli editoriali.

¹⁸ "Nature" 'Open access' will not be open to everyone. Making authors pay for publication may not deliver the anticipated benefits. Vol 425 No 6958 p. 559 (9 Oct 2003)

A more extensive commentary entitled "The Public Library and the Public Good" of the Ewing's letter it's freely available at

<http://www.ams.org/ewing/Documents/PublicLibrary-70.pdf>

Chiudo questa *release* scritta della nostra performance orale, in omaggio a Luisella Goldschmidt-Clermont che con il suo lavoro pionieristico sui pre-print ha segnato una svolta epocale nei circuiti della comunicazione scientifica.

*“Una fitta folla circonda un palco: qualche sgomitata sembra essere messa in relazione al processo filtrante attraverso il quale le persone sono ammesse una alla volta a salire sul palco, scarabocchiare sulla lavagna e parlare. Prima di andar via scambiano occasionalmente alcune frasi con qualcuno del pubblico. A parte questi contributi minori, il pubblico è generalmente tranquillo, rispettoso se qualcuno fra di loro dorme. Ma appena l'ultimo oratore chiude lo spettacolo abbandonando il palco, la folla si anima: si formano sottogruppi, conversazioni, argomenti vengono fuori da ogni dove...”*¹⁹

Luisella aveva ragione, le conferenze sono una parte essenziale di questo processo perché forniscono la grande occasione di scambi di idee a beneficio della comunità.

¹⁹ Goldschmidt-Clermon, Luisella (1965), op. cit.