

Open Archive per una comunicazione scientifica "free online"

di Antonella De Robbio
Referente SBA per il diritto d'autore
Università degli Studi di Padova

"Liberare la ricerca valutata, presente e futura,
da tutti gli ostacoli all'accesso e all'impatto,
dipende ormai dai ricercatori.
La posterità già ci guarda e non ci giudicherà bene
se continuiamo a ritardare senza ragione
l'ottimo e l'ineluttabile,
che sono chiaramente alla nostra portata".

[Stevan Harnad¹]

Abstract

Tutta la ricerca dovrebbe passare attraverso i canali dell'accesso libero, perché nessuno paga gli autori per i loro lavori scientifici. Le produzioni intellettuali dei ricercatori dovrebbero essere, secondo molte scuole di pensiero e anche secondo chi scrive, completamente sganciate dai meccanismi di mercato validi per gli autori che ricevono "royalties" dagli editori (letteratura non scientifica). Per questo parliamo di modelli "aperti" di comunicazione scientifica "free online".

Questi modelli poggiano su software libero e, come dice Richard Stallman, fondatore della FSF Free Software Foundation, attraverso il 'copyleft', giuridicamente basato sul copyright, si consente una piena e totale libertà di utilizzo da parte degli utenti.

Il presente lavoro è allineato con le teorie di Stevan Harnad², e con la filosofia su cui poggia il movimento FOS Free Online Scholarship coordinato da Peter Suber³, entrambi aderenti alle politiche di Open Archive Initiative⁴.

¹ Stevan Harnad, *For Whom the Gate Tolls? How and Why to Free the Refereed Research Literature Online Through Author/Institution Self-Archiving, Now*. Department of Electronics and Computer Science, University of Southampton
<<http://www.cogsci.soton.ac.uk/~harnad/Tp/resolution.htm>>.

² Vedi nota precedente, <<http://www.cogsci.soton.ac.uk/~harnad/>>

³ Per le linee politiche del movimento FOS cfr. Editorial Position of the FOS Newsletter, version 1.6, <<http://www.earlham.edu/~peters/fos/index.htm>>.

Gli ostacoli al diritto di accesso ai contenuti fanno capo a due paradossi. I prezzi di abbonamento alle riviste, sempre più alti, che rendono inaccessibili i contenuti, e i diritti di proprietà intellettuale "degli autori". Questi ultimi finiscono gratuitamente nelle mani degli "editori" i quali, invocando la legge, non consentono la duplicazione dei contenuti. Le regole del gioco vanno cambiate. L'auto-archiviazione dei lavori scientifici entro gli archivi aperti (Open Archive) istituzionali o disciplinari è un passo ormai non più procrastinabile.

Letteratura scientifica "free online"

Numerosi sono gli studi che si stanno occupando di analizzare i meccanismi che regolano il circuito della comunicazione scientifica. Altrettanto numerosa la letteratura proveniente dalla comunità degli stessi scienziati, che sta mettendo a fuoco una situazione ormai insostenibile nel tentativo di stravolgere le parti. Alcune scuole di pensiero stanno proponendo valide soluzioni alternative all'aumento dei prezzi delle riviste, cosa che provoca considerevoli danni alla ricerca, in quanto oltre alla sottrazione di denaro pubblico alla comunità dei ricercatori, comporta una riduzione dell'impatto dei lavori all'interno della comunità stessa. Tali soluzioni in certi ambienti scientifici sono allo stato di sperimentazione o in fase di decollo, in altri sono esperienze già consolidate da anni.

La delimitazione delle produzioni intellettuali, entro territori "controllati" da terze figure del mondo commerciale, editori, produttori e aggregatori, se da una parte è garanzia di "qualità", dall'altra risulta essere un blocco alla crescita scientifico-culturale.

Stevan Harnad, studioso nel campo delle scienze cognitive e ideatore del sito "eprints"⁵, parla di "una linea di demarcazione profonda ed essenziale nella galassia Post-Gutenberg" tra le opere che rientrano nel circuito commerciale - in tutte le fasi che coinvolgono l'autore nella cessione dei relativi diritti dietro compenso (royalty) ed essenzialmente accessibili a pagamento (libri, giornali, programmi, musica) - e le opere che rientrano nel circuito della comunicazione scientifica, il cui esempio più rappresentativo è quello degli articoli di ricerca pubblicati nelle riviste scientifiche.

Gli ostacoli ad un libero accesso ai contenuti sono ostacoli all'impatto scientifico dei lavori entro la comunità e, a lungo andare entro l'intera collettività, e conducono ad un rallentamento nella crescita della ricerca scientifica e del progresso tecnologico.

Laddove una rivista scientifica pubblicata da un editore distribuisce i suoi contenuti a prezzi elevati o a condizioni di accesso restrittive, prima saranno i singoli studiosi della materia a rinunciare ad abbonarsi, ma via via anche le biblioteche saranno costrette a disdire l'abbonamento (su carta, ma anche online), penalizzando intere fasce di utenti.

Una rivista scarsamente presente nelle biblioteche perché costosa, non sarà aperta o consultata da nessuno e nessuno potrà mai leggerne i contenuti. Ne consegue una perdita di efficacia dei contenuti pubblicati in quella rivista, a causa di un sistema di distribuzione inefficiente che non consente una fruizione in termini di efficacia.

⁴ <<http://www.openarchives.org/>>

⁵ <<http://www.eprint.org>> mette a disposizione software per la creazione di data server OpenArchive Initiative compatibili

Se infatti andiamo ad analizzare le due coordinate costituite dal "fattore di impatto" e dal "prezzo della rivista", potremo avere notevoli delle sorprese.

Molti sono i movimenti "di libertà" che stanno dettando nuove regole entro un mercato che poggia sugli "archivi aperti", noti con il termine di "Open Archive", server-contenitori organizzati dove è possibile depositare liberamente i propri lavori scientifici. Questi modelli, ancora in fase embrionale ma già ben ancorati in un'architettura consolidata entro la rete, per certe discipline sono già in fase avanzata, sebbene vi siano state esperienze passate di difficile applicabilità che hanno condotto oggi a ragionare in termini di "interoperabilità" come requisito fondamentale all'apertura di un server. Queste nuove regole sono codificate in protocolli che parlano linguaggi aperti.

Stevan Harnad in testa - e in buona compagnia di eminenti economisti - dimostra come le questioni connesse al diritto d'autore o copyright incidano sui processi che caratterizzano il circuito della comunicazione scientifica.

La questione della proprietà intellettuale sui lavori scientifici ha un'influenza diretta sull'andamento dei prezzi delle riviste. Il paradosso nasce dal fatto che la normativa sulla proprietà intellettuale non differenzia i due canali, quello del circuito della ricerca da quello commerciale.

Tracciare una doverosa linea di demarcazione che separa due mondi differenti (scientifico e commerciale) conduce necessariamente a riequilibrare le parti in causa e a ripristinare in modo equo ciò che spetta a ciascuna delle due sfere. Vi è infatti una notevole differenza tra quello che l'autore cede a seguito di compenso (*royalty*) e quello che l'autore cede gratuitamente (*give away*). Non è la stessa cosa scrivere un articolo giornalistico, o per una rivista popolare, e scrivere un articolo scientifico di ricerca. Meccanismi diversi vanno regolati da leggi differenti.

Alcuni editori consentono di collocare i lavori sui siti personali Web degli autori, ma non sugli eprints server, mentre altri permettono agli autori di depositare la versione pre-print di un lavoro, ma non quella finale pubblicata dall'editore.

Ha fatto molto scalpore a livello internazionale, con larga eco nelle liste di discussione che gravitano attorno ai server di eprints, l'ambigua frase presente nell'intervista rilasciata da Derk Haank di Elsevier e pubblicata su "Information Today" dell'aprile 2002⁶: "You can put your paper on your own Web site if you want. I believe we have the most liberal copyright policy available".

Definiamo ambigua la frase perché, come rilevato dai carteggi della lista di discussione "cogprints", i contratti sottoscritti dagli autori con Elsevier riportano:

Posting of a preprint version of this work on an electronic public server is permitted. Posting of the published article on a secure network (not accessible to the public) within the author's institution is permitted. However, posting of the published article on an electronic public server can only be done with Elsevier's written permission.

⁶ Richard Poynder, *Not Pleading Poverty Interview of Elsevier*. "Information Today", 19 (2002) 4, <<http://www.infotoday.com/it/apr02/poynder.htm>>.

Quest'ultima frase sembra molto diversa da quella dell'intervista, dove si afferma che l'importante è non cedere ad altri editori, per scopi di lucro, lo stesso lavoro ceduto a Elsevier. Nella clausola contrattuale sopra riportata sembra che sia possibile depositare sul server istituzionale ad accesso pubblico il pre-print, ma non la versione "pubblicata". Che cosa si intende per versione pubblicata non è dato a sapersi: e se il pre-print restasse immutato, senza variazioni rispetto alla versione pubblicata? O si intende piuttosto il PDF con il layout dell'editore? Parliamo di forme o di contenuti? Di espressioni o di manifestazioni (per usare una terminologia da bibliotecari)?

In ogni caso è possibile depositare anche l'articolo pubblicato a patto che sia posto in luogo protetto, per esempio in un server che limiti l'accesso ad un determinato gruppo di utenti e solo a quelli e, a quanto pare, a utenti interni la rete locale.

Ci sono diversi modi di intendere un archivio aperto. Quello di Elsevier è un modo ancora legato ad un concetto di "distribuzione" che ruota attorno al modello della pubblicazione cartacea. Il nostro invece è un modello che porta dentro il "seme" della parola "disseminazione". Disseminare conoscenza per raccogliere ulteriore conoscenza. E' un concetto profondamente diverso da quello del "distribuire", che implica un'offerta, gratuita o a pagamento, ma sempre senza ritorno, laddove per ritorno si intende un ritorno in termini di conoscenza e non puramente economico. La disseminazione implica una crescita. Per questo, quando parliamo di archivi aperti, parliamo sempre di disseminazione della letteratura scientifica, quando invece ci riferiamo al contesto commerciale a pagamento, parliamo in termini di distribuzione.

Elsevier ha messo a disposizione più server per il deposito dei pre-print scientifici: uno per la matematica, uno per la chimica e uno per l'informatica. Quello relativo alla matematica è componente di un portale denominato "Math Portal", il quale offre un servizio di ricerca entro le pubblicazioni Elsevier e di altri editori consorziati attraverso l'utilizzo di metadati.

Chiunque può depositare su questi tre server, che sono stati i primi ad essere nati sotto la diretta protezione di un editore commerciale. Gli autori che invece hanno pubblicato e ceduto i diritti a Elsevier, a detta di Micheil Kolman, direttore editoriale di Elsevier per la matematica e l'informatica, possono depositare anche la versione definitiva pubblicata sul server di Elsevier a disposizione di tutti liberamente.

Senza nulla togliere all'importanza dell'iniziativa, si tratta inequivocabilmente di un'ovvia manovra pubblicitaria, dal momento che articoli liberamente accessibili sul server di pre-print e collocati dagli stessi autori fanno da calamita al servizio commerciale di ScienceDirect. Il server si distingue per la sua configurazione inusuale: infatti, per il recupero dei documenti, utilizza un sistema di ranking per stabilire la rilevanza dei risultati ottenuti.

Open Archive per una disseminazione scientifica aperta

L'avvento dei server di preprint, soprattutto in certe discipline come la matematica, è incrementato a tal punto da evolvere e divenire una forma di comunicazione e di immagazzinamento della produzione scientifica nella sua interezza.

Molti sono i ricercatori che da alcuni anni archiviano i loro lavori sui server delle loro istituzioni o in server disciplinari, per esempio per la biomedicina, l'astronomia, la chimica.⁷ I fisici già da dieci anni archiviano i loro lavori nei server di tutto il mondo, in quanto furono i primi a partire con il server xxx.lanl.gov, ora arXiv.org, modello di server centralizzato che per anni è stato sovrano in questo campo.

Anche i matematici hanno sviluppato progetti internazionali per la creazione di *repositories* dei loro lavori⁸. Tali progetti sono per esempio *Front for the Mathematics*, con oltre 10.000 lavori⁹, e *ArXiv*¹⁰ entro il server dei fisici di Los Alamos. Il progetto europeo *MPRESS/MathNet.preprints server*¹¹, coordinato dai matematici tedeschi Judith Plümer, Roland Schwänzl, non è un server per il deposito, ma un'architettura di tipo "umbrella server", la quale consente di ricercare nei metadati di alcuni server di preprint, attraverso una meta-interfaccia. Infine i ricercatori in scienze cognitive in tre anni hanno depositato oltre mille articoli¹².

Se teniamo da parte la fisica, la matematica è il campo in cui si è fatto l'uso più massiccio di server di pre-print. Ciò è chiaramente dovuto all'influenza di arXiv.org di Los Alamos il quale, a partire dagli anni novanta, ha avuto un ruolo strategico, per esempio durante il periodo della crescita della teoria delle super stringhe, o per i settori dell'algebra e della geometria differenziale, a tal punto da svolgere funzione di *clearinghouse* per lavori in campi in rapido movimento.

Ma anche all'interno di una stessa disciplina non sempre i comportamenti sono omogenei. Non tutti i matematici utilizzano i server per il deposito di materiale scientifico. Per esempio l'analisi numerica è un settore con tradizioni stabilizzate da molto tempo nell'uso di strumenti di comunicazione tipici dell'era pre-Web, dove la comunicazione avviene essenzialmente via e-mail e le produzioni scientifiche scambiate entro mailing list o newsgroup.

Stranamente anche l'informatica, ad eccezione dell'esperienza di NCSTRL, Networked Computer Science Technical Reference Library¹³, difetta dell'uso dei preprint server, dato che è invalsa piuttosto l'abitudine di collocare i propri lavori su siti e pagine Web personali. Questo modo di procedere crea una situazione di tipo frammentato difficilmente gestibile,

⁷ Antonella De Robbio, *Mathematics and Physics Preprint and e-Print Servers* <<http://library.cern.ch/derobbio/mathres/preprint.html>>.

⁸ <<http://www.ams.org/global-preprints/umbrella-server.html>>.

⁹ Front End for the Mathematics ArXiv, <<http://front.math.ucdavis.edu/>>.

¹⁰ <<http://arXiv.org/>>.

¹¹ MPRESS - MathNet.preprints, <<http://MathNet.preprints.org/>>.

¹² CogPrints Archive The Cognitive Sciences E-Print Archive (Computer Science and Engineering, Psychology, Neuroscience, Behavioral Biology, Linguistics and Philosophy)⁹, <<http://cogprints.soton.ac.uk/>>.

¹³ <<http://www.ncstrl.org/>>.

ragione per cui, potremmo forse arguire maliziosamente, certi editori acconsentono ad una collocazione dei papers su siti Web personali, ma non il deposito dentro un archivio aperto organizzato.

Per sanare questa frammentarietà è nato ResearchIndex, servizio precedentemente chiamato CiteSeer, costruito da NEC Research Institute il quale pesca regolarmente home page degli studiosi nel campo dell'informatica, catturando e collezionando papers. ResearchIndex che consente ai singoli autori di depositare i propri lavori è peraltro connesso con arXiv attraverso una sezione relativa all'informatica, chiamata CoRR Computing Research Repository. CoRR è un modello ibrido tra arXiv e NCSTRL, nato nel 1997 in seno ACM Association for Computing Machinery. La differenza tra il deposito in un server, e il collocare pagine in Web catturabili da ResearchIndex è che in CoRR vi è una certa garanzia di stabilità e permanenza di quanto depositato.

Non parlerò in questa sede di NCSTRL, che si pronuncia "ancestral", in quanto esiste un piano di conservazione e transizione dal vecchio modello che implementava il protocollo Dienst verso il modello che utilizza tecnologie Open Archive Initiative. La trasformazione di NCSTRL coinvolgerà probabilmente anche ERCIM, The European Research Consortium for Informatics and Mathematics, che ne è parte integrante.

Mentre per la matematica, la fisica, l'informatica l'uso pervasivo del linguaggio TeX ha comportato una rivoluzione nel creare documenti disponibili liberamente in formato elettronico, in altre discipline come la medicina, la biologia e o la chimica gli open archive trovano ostacoli di varia natura. David Morrison della Dyke University e responsabile della sezione matematica di arXiv afferma che i matematici hanno uno "spirito open source" forse ereditato geneticamente dallo stesso Donald Knuth, fondatore del TeX.

Negli Stati Uniti nell'ultimo decennio sono sorti numerosi archivi aperti di varia natura e/o tipologia, in Europa vi sono alcune esperienze assai notevoli e anche in Italia stanno sorgendo i primi agglomerati per certi settori disciplinari o basati su server istituzionali.

Esistono modelli centralizzati e modelli decentrati, o sistemi a modello distribuito o archivi decentralizzati.

Tipico esempio di modello centralizzato è arXiv.org, dove peraltro sono confluiti, nel corso degli ultimi anni, piccoli server disciplinari. Difatti il dibattito recente imperversa anche attorno alla questione "big versus small"¹⁴.

Attualmente si stima che ArXiv serva circa 80.000 utenti di oltre 70 Paesi. Nel 2000 vi sono stati 13 milioni di documenti scaricati. La crescita delle sottomissioni è pari a 3500 nuovi lavori annui, per oltre 33.000 sottomissioni a fine 2001.

La centralizzazione di ArXiv è stato un processo che è iniziato a partire dal 1994, poiché prima esistevano piccoli archivi su base disciplinare che funzionavano con lo stesso software.

¹⁴ Allyn Jackson, *From Preprint to E-prints: the rise of electronic preprint servers in mathematics*. "Notices of the AMS", January 2002, p. 23-31.

Nel mondo sono centinaia i server di preprint, e se dobbiamo fare una distinzione basata sul livello disciplinare, questi si possono suddividere in due ampie categorie: i piccoli server disciplinari settoriali (*subject-based server*) e i server generali che, pur essendo relativi a una disciplina, sono server istituzionali di istituti o singoli dipartimenti.

La distinzione tra server disciplinare settoriale e server generale si individua entro la stessa disciplina, ma se parliamo di server istituzionali multidisciplinari ci troviamo di fronte ad una categoria a parte che è quella dei "server istituzionali centralizzati", i quali ospitano i lavori di tutta la comunità che afferisce a quell'istituzione (università, ente di ricerca, ...).

Questi server esplicano una funzione di deposito per i lavori di ricerca che si producono entro l'istituzione, e possono risultare utili momenti di raccordo con i nuclei di valutazione della ricerca di un paese. Nel nostro paese è giunto il momento di istituire queste "zone accademiche free online", opportunamente coordinate attraverso uno schema gerarchico multi-disciplinare adeguato alla figura dell'utente-autore il quale utilizza le nostre biblioteche accademiche e di ricerca.

Martha Toker, bibliotecaria alla Washington University, osserva che un grande server centralizzato è molto migliore di tanti piccoli server diversi per tipologia e localizzazione. "Bigger is better", si sostiene da molte parti confrontando i costi del personale, che sono molto maggiori rispetto ai costi di immagazzinamento dei papers. Monitorare costellazioni di server potrebbe essere molto più laborioso che gestire un unico grande server.

Le esperienze di "umbrella server" del tipo MPRESS offrono un utile servizio di meta-ricerca che consente di interrogare più server attraverso metadati Dublin Core, ma tale soluzione non offre un deposito unico.

Hans Koch invece sostiene la teoria opposta, in quanto è fermamente convinto che un sistema distribuito sia senza dubbio migliore, lasciando peraltro inalterata la visione di un grande server unico virtuale realizzato attraverso la raccolta e l'indicizzazione dei metadati a livello centrale, mentre il posseduto reale rimarrebbe distribuito a livello locale.

Un vero modello di sistema di archivi aperti di tipo distribuito è RePEc, la più grande collezione decentrata di documenti ad accesso libero per l'economia, trattandosi di una rete che collega diversi archivi. Il progetto WoPEc (Working Papers in Economics), sorto nel 1999 per opera di Thomas Krichel¹⁵, è costituito dalla rete RePEc, la quale connette oltre 200 archivi¹⁶ che mettono a disposizione oltre 60.000 articoli e rapporti tecnici di ambito economico disponibili online. Ogni istituzione che partecipa al progetto mette a disposizione i propri dati ma, contemporaneamente, diventa un *mirror* che garantisce l'accesso anche ai dati di tutti gli altri archivi.

Probabilmente entrambe le soluzioni, con tutte le eventuali sfumature e anche le sovrapposizioni, dovrebbero convivere in quanto sono entrambe necessarie, a seconda delle

¹⁵ Economista ora docente di biblioteconomia a New York.

¹⁶ Per la lista degli archivi si veda <<http://ideas.uqam.ca/ideas/archives.html>>.

situazioni, della disciplina, e di tante altre variabili. "Non esiste un'unica grande biblioteca fisica centrale sulla terra, ma piuttosto esistono migliaia e migliaia di biblioteche"¹⁷. La stessa cosa vale per i gli archivi aperti.

Per questa ragione gli archivi elettronici nella piattaforma Open Archive Initiative (OAI) sono tutti concepiti in interoperabilità gli uni con gli altri e funzionanti con software libero.

Gli archivi aperti possono essere resi ricercabili come se facessero parte di un archivio virtuale globale con tutta la letteratura scientifica disponibile, pre e post valutazione.

A tal fine ogni nuova installazione deve essere opportunamente registrata nell'apposito registro internazionale presso il sito OAI¹⁸, iniziativa sorta a seguito del numero crescente di preprint elettronici e di conseguenza di archivi di preprint disponibili in rete, diversi per struttura, per metadati e differenti anche a livello organizzativo.

L'Open Archive Initiative nasce in ambito statunitense nel settembre 2000 e si concretizza già nel successivo mese di ottobre al meeting di Santa Fe, nel Nuovo Messico, intorno alle idee di Paul Ginsparg (Los Alamos National Laboratories), Herbert Van de Sompel e Rick Luce (Cornell University).

Il primo incontro, noto come meeting di Santa Fe, fu organizzato dal Council on Library and Information Resources (CLIR), DLF, The Scholarly Publishing and Academic Resources Coalition, The Association of Research Libraries, e The Research Library of the Los Alamos National Laboratory, avendo come obiettivo l'interoperabilità tra archivi di preprint.

OAI fonda le sue radici nello sforzo continuo per il miglioramento dell'accesso ai documenti negli archivi e-print, attraverso la messa a disposizione strumenti, software e documentazione (protocolli, standard) e la promozione di metadati standard.

Mentre molti di questi archivi iniziarono la loro esistenza come veicoli informali per la disseminazione dei risultati preliminari delle ricerche all'interno di specifiche comunità di studiosi, alcuni server cominciarono via via ad evolversi all'interno di un nuovo medium, strategico e sempre più essenziale, che consentiva la condivisione dei risultati delle ricerche ai colleghi che operavano nello stesso campo di indagine.

Se i primi modelli essenzialmente potevano configurarsi come contenitori di materiale *non-peer-reviewed*, chiamato anche *gray literature*, i modelli successivi andavano sempre più conformandosi all'interno di processi più complessi che interagivano con le altre metarisorsse (OPAC, banche dati, ...) grazie all'azione delle biblioteche.

In OAI esistono due categorie di server:

- i data provider, i quali contengono i documenti depositati e che supportano il protocollo OAI esponendo i metadati relativi ai lavori contenuti nel server; un data provider gestisce uno o più server, per il deposito (Web servers repositories)

¹⁷ Peter Michor, in Allyn Jackson, *From Preprint to E-prints: the rise of electronic preprint servers in mathematics*. "Notices of the AMS", January 2002, cit, p. 29.

¹⁸ <<http://www.openarchives.org/data/registerasprovider.html>>.

- i service provider, che usano i metadati dei data provider come base per la costruzione di servizi a valore aggiunto.

Ricerca preprint, o comunque reperire fonti e papers entro queste zone grigie, è un compito assai arduo e molto lavoro vi è ancora da fare, soprattutto a livello di servizi che conducano, attraverso i service provider preposti alla raccolta di metadati, alla costruzione di un archivio virtuale globale come auspicato da Harnad.

CiteBase ad esempio è parte di uno sforzo teso a migliorare i servizi online per la comunità scientifica: mette a disposizione il software per la costruzione di archivi (Eprints), il software per il parsing citazionale (OpCit) ai fini della connessione tra citazioni (reference linking). I servizi Open Archive di CiteBase sono una ricca fonte di informazione e un punto di partenza verso il movimento dell'auto-archiviazione.

CiteBase Search¹⁹, disegnato e scritto da Tim Brody, fa parte del progetto Open Citation Project²⁰, fornisce agli utenti la possibilità di ricercare entro archivi multipli differenti con funzionalità di ranking nella presentazione dei risultati, incluso anche l'impatto della citazione. I dati sono raccolti usando il protocollo Open Archives Metadata e un harvester che si occupa di scaricare i dati da un archivio all'altro, in questo caso da ArXiv²¹ e da Cogprints²², l'archivio aperto per le scienze cognitive.

Fondamentale è l'organizzazione di service provider dedicati alla raccolta di metadati dai vari archivi, o sistemi di *metadata harvesting*²³, entro una cornice di interoperabilità estesa al di là delle differenze tra server. Va detto che, nell'implementare un server OAI, sarebbe opportuno attenersi al protocollo Open Archives Initiative Protocol for Metadata Harvesting Protocol²⁴, indispensabile per rendere visibili ai fini della raccolta, i propri metadati esposti, in un colloquio entro la comunità internazionale.

¹⁹ <<http://citebase.eprints.org/>>.

²⁰ Al progetto OpCit prendono parte Stevan Harnad, Les Carr, Zhuoan Jiao, Steve Hitchcock dell'Intelligence, Agents and Multimedia Group presso L'Università di Southampton

²¹ Archivio gestito da Paul Ginsparg e Simeon Warner

²² Archivio ideato e diretto da Stevan Harnad e gestito da Chris Gutteridge

²³ Per esempio ALCME: OAIHarvester Project, <<http://alcme.oclc.org/OAIHarvester.html>>.

²⁴ The Open Archives Initiative Protocol for Metadata Harvesting
Protocol Version 2.0 of 2002-06-14. Document Version 2002/07/05T16:47:00Z
><http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>>

Cross Archive Searching Service²⁵ è un altro *harvester* che consente di ricercare in tutti gli Open Archives; funziona con Arc²⁶, il quale raccoglie i metadati da diversi archivi OAI compatibili, li normalizza e li immagazzina in un servizio di ricerca basato su un database relazionale del tipo MySQL o Oracle. Ad oggi, in Arc vi sono oltre 320.000 metadati raccolti da 18 data provider provenienti da vari domini di soggetto. E' stato anche implementato un puntatore OAI su Arc, al fine di ottenere una raccolta gerarchica dei metadati. Si tratta della prima notevole esperienza di servizio di ricerca cooperativo²⁷ basato sul protocollo OAI, originato dal precedente prototipo Universal Preprint Service (UPS)²⁸, che è servito da base per Arc. Tale esperienza dovrebbe essere presa in seria considerazione da quanti vogliono costruire *service provider* simili ad OAI.

Interessante è anche il progetto UIUC OAI Metadata Harvesting Project dell'University of Illinois Urbana-Champaign²⁹, entro il quale sono stati sviluppati numerosi strumenti, con software a codice sorgente aperto³⁰, oltre ad una meta-interfaccia di ricerca Search Harvested Metadata, che offre accesso alle collezioni del loro patrimonio.

Il progetto OAIster³¹ invece, che fa parte del servizio di produzione della biblioteca digitale dell'Università del Michigan, adotta l'iniziativa Metadata Harvesting della Mellon Foundation. L'obiettivo è rendere ricercabili un gran numero di archivi per il deposito (*repositories*), liberamente accessibili e utili alla ricerca.

TORII³² è il progetto della SISSA di Trieste per un accesso unificato ai vari archivi aperti per la fisica e l'informatica.

²⁵ <<http://arc.cs.odu.edu/>>.

²⁶ Xiaoming Liu, Kurt Maly, Mohammad Zubair Michael L. Nelson, *Arc - An OAI Service Provider for Digital Library Federation*. "D-Lib Magazine", 7 (2001) 4, <<http://www.dlib.org/dlib/april01/liu/04liu.html>>.

²⁷ In inglese "federated".

²⁸ Van de Sompel, Herbert, Thomas Krichel, Micheal L. Nelson, et al. (2000). *The UPS Prototype project: exploring the obstacles in creating a ross eprint archive end-user service*). D-Lib Magazine, Feb. 2000, 6(2) <<http://www.dlib.org/dlib/february00/vandesompel-ups/02vandesompel-ups.html>>

²⁹ <<http://sourceforge.net/projects/uilib-oai/>>.

³⁰ Gli strumenti offerti sono: Metadata Stored in Microsoft Access, Metadata Stored in XML files, Metadata Maintained in HTML files, <<http://oai.grainger.uiuc.edu/ProviderTools/>>.

³¹ <<http://oaister.umdl.umich.edu>>.

Anche NCSTRL³³ fornisce un accesso unificato ai rapporti tecnici e preprints di informatica provenienti da dipartimenti e laboratori della rete NCSTRL. Si tratta di un'implementazione OAI che utilizza il protocollo *metadata harvesting*, il quale ha sostituito la vecchia architettura Dienst, molto complessa da gestire e che richiedeva personale altamente specializzato a livello informatico.

La Budapest Open Access Initiative (BOAI) e il movimento Free Online Scholarship (FOS)

In tale ottica è importante segnalare alcune iniziative, suddivise in due aree distinte ma strettamente in relazione.

Difatti da una parte si situa l'Open Archive Initiative (OAI) descritta sopra, la quale è nata con lo scopo di sviluppare e promuovere standard per l'interoperabilità tra archivi al fine di essere di supporto ad un'efficiente disseminazione di contenuti; dall'altra si collocano iniziative come la Budapest Open Access Initiative (BOAI) o il movimento Free Online Scholarship (FOS).

L'iniziativa di Budapest si distingue da OAI dal momento che il suo scopo è quello di fare pressione sugli autori al fine di ottenere un consenso generalizzato. Per BOAI ciò che conta è il consenso dell'autore, non soltanto le necessità o i desideri dell'utenza. BOAI non si occupa di applicazioni software o di aspetti tecnologici, ma è focalizzata principalmente sui processi di peer-review.

Il suo scopo primario è quello di promuovere l'auto-archiviazione. Il concetto che sottende la filosofia BOAI è quello di un rigoroso rispetto del diritto d'autore, che libera la ricerca. E' un concetto molto vicino a quello del software libero distribuito con licenza GNU, di cui parlerò in seguito.

Dal 14 febbraio ad oggi (10 luglio 2002), in cinque mesi sono 2473 i ricercatori, bibliotecari o altre persone che ruotano attorno al mondo dell'informazione che hanno sottoscritto l'iniziativa³⁴, finanziata dalla Fondazione Soros di OSI Open Society Institute. OSI elargirà un milione di dollari per anno, per tre anni consecutivi, per lo sviluppo di un progetto il quale prevede piani e modelli per l'auto-archiviazione "sostenibile" in oltre 140 Paesi. OSI è nata nel 1993 dal sostegno finanziario di Soros, allo scopo di offrire supporto allo sviluppo di programmi (sociali, educativi, per i diritti delle donne, ...) nei Paesi in via di sviluppo: Europa dell'Est, ex Unione Sovietica, Haiti, Mongolia, Africa, America Latina, etc.

Tra gli ideatori di BOAI ci sono nomi come Fred Friend (Director Scholarly Communication, University College London), o Rick Johnson (Director, Scholarly Publishing and Academic Resources Coalition (SPARC) e anche rappresentanti del mondo dell'editoria, Jan Velterop di BioMed Central per esempio.

³² <<http://tips.sissa.it>>.

³³ <<http://www.ncstrl.org/>>.

³⁴ <<http://www.soros.org/openaccess/index.shtml>>.

Il movimento Free Online Scholarship (FOS) coordinato da Peter Suber ha una posizione editoriale la quale prevede le seguenti linee:

- ❑ la letteratura scientifica dovrebbe essere libera e online
- ❑ letteratura scientifica per FOS assume il significato di letteratura professionale di ricerca in ogni campo scientifico e umanistico
- ❑ il significato del termine "libero" per FOS significa libertà da pagamenti per l'accesso, ma anche assenza di ogni censura
- ❑ "online" per FOS significa disponibilità dei documenti in Internet in una forma in cui il testo pieno possa essere letto, copiato, connesso, stampato, scaricato e ricercabile
- ❑ la letteratura scientifica "free online" può assumere molte forme. Può consistere di pre-print o post-print, articoli o volumi o presentazioni multimediali. Può soppiantare o essere supplementare a edizioni a stampa. Può utilizzare processi di peer-review retroattivi o futuri o potenziali. Può essere presente fuori da contenitori organizzati, in un sito web, o risiedere all'interno di un archivio organizzato a livello istituzionale, o disciplinare, o entro un periodico elettronico, o entro una conferenza, o comunque organizzato in altro modo. Ci sono anche altre varianti, tenendo conto che esistono anche quelle non completamente libere e non completamente online. La coesistenza e cooperazione di molte di queste forme sarà pertanto necessaria
- ❑ Il processo di peer-review può avvenire efficacemente sia per le pubblicazioni a stampa, sia per le pubblicazioni in linea. Ci sono infiniti modi per facilitare forme di peer-review entro un ambiente elettronico interattivo. FOS incoraggia sperimentazioni con monitoraggio dei risultati
- ❑ la stampa non è un nemico. Se lo stesso documento è pubblicato sia online che a stampa tanto meglio, a meno che il costo dell'edizione a stampa non conduca l'editore a caricare l'edizione online o a negare il permesso agli autori di mettere i propri lavori online
- ❑ il copyright non è un nemico. La letteratura protetta da copyright può essere pienamente e liberamente accessibile. FOS potrebbe anche utilizzare il copyright per prevenire forme di sfruttamento commerciale. Ciononostante FOS potrà accrescersi più velocemente se vi sarà un modo intelligente di revisionare le leggi sulla proprietà intellettuale per l'ambito scientifico e se i giornali adotteranno la pratica di lasciare il copyright nelle mani degli autori
- ❑ il consenso dell'autore è condizione necessaria. L'unica letteratura scientifica che dovrebbe essere libera è quella i cui relativi autori desiderano che sia libera
- ❑ "call for free" per la letteratura scientifica è in primo luogo un appello agli autori a mettere i loro lavori online, volontariamente e liberamente accessibili. In secondo luogo è un appello agli editori, invitandoli a mettere le loro pubblicazioni online, volontariamente e liberamente accessibili. Questo non è certamente un appello alla pirateria o un obbligo a rendere libera e aperta la letteratura "chiusa" contro la volontà degli autori. D'altro canto, mentre gli autori e gli editori hanno diritto di stabilire il prezzo sulla loro letteratura, la difesa del prezzo delle copie delle opere dovrebbe essere compatibile con i diritti dei lettori e comunque tener conto delle eccezioni per la reprografia per equi utilizzi, per le copie di back-up e così via
- ❑ la letteratura "free online" non è senza problemi. Uno dei più importanti è quello legato alla conservazione e all'accesso a lungo termine. Un altro problema è persuadere gli

autori, ma soprattutto le commissioni di concorso, a dare lo stesso peso alla valutazione dei periodici sia a stampa sia elettronici

- differenti discipline hanno diverse necessità.

Il concetto "Open Source" nel software libero

Richard Stallman, fondatore della FSF *Free Software Foundation*³⁵, riferendosi al mondo del software libero, spiega che il 'copyleft' riesce a rendere operativo l'ideale cui si ispira, ossia la libertà degli utenti, perché è giuridicamente basato sul copyright.

L'espressione "software libero"³⁶ secondo la filosofia GNU³⁷, un sistema operativo che è interamente costituito da software libero, si riferisce alla libertà dell'utente di eseguire, copiare, distribuire, studiare, cambiare e migliorare il software. Più precisamente, esso si riferisce a quattro tipi di libertà per gli utenti del software:

1. l'utente ha la libertà di eseguire il programma per qualsiasi scopo
2. l'utente ha la libertà di modificare il programma secondo i propri bisogni (perché questa libertà abbia qualche effetto in pratica, è necessario avere accesso al codice sorgente del programma, poiché apportare modifiche ad un programma senza disporre del codice sorgente è estremamente difficile)
3. l'utente ha la libertà di distribuire copie del programma, gratuitamente o dietro compenso
4. l'utente ha la libertà di distribuire versioni modificate del programma, così che la comunità possa fruire dei miglioramenti apportati.

Il termine "Open Source" nel caso del software è usato con lo stesso significato di software libero, in un'ampia accezione di libertà.

I software liberi o Open Source sono rilasciati sotto una licenza nota come GNU/GPL General Public License, definita da Richard Stallman che ha ideato tale licenza.

Nel software libero la tecnica del 'copyleft' consiste nel dare il permesso, tramite la licenza, di modificare il programma, di distribuirlo, e di pubblicarne una versione perfezionata entro dei termini che devono essere sempre gli stessi, di modo che chiunque ottiene una copia del software ottiene la stessa libertà che il primo utente ha avuto.

La licenza GPL, geniale nella sua concezione, impone la distribuzione del codice sorgente assieme ai programmi e una distribuzione dei software basati sui quei codici sorgenti che devono anch'essi essere distribuiti in licenza GPL. "Se si rendesse il software di dominio pubblico permetteremmo alle varie società poco scrupolose - che producono software non

³⁵ <<http://www.free-soft.org/>>.

³⁶ Il termine *free* in inglese significa sia gratuito che libero.

³⁷ Richard Stallman, *Linux ed il Progetto GNU*, <<http://www.gnu.org/gnu/linux-and-gnu.it.html>>.

libero - di prendere i nostri programmi e farne versioni modificate e di distribuirle come software 'proprietario' senza alcuna libertà".

Il refereeing nel circuito della comunicazione scientifica

Aspetto centrale che attiene la sfera della valutazione è il refereeing degli articoli, in cui esperti di un settore disciplinare controllano e certificano la qualità del lavoro di colleghi anch'essi esperti. Senza il controllo di qualità, la letteratura scientifica non sarebbe né affidabile, né navigabile all'interno del meccanismo delle citazioni bibliografiche tra referenze.

La certificazione di qualità, nota col la sigla QC/C *quality control and certification*, a seguito dell'attività di refereeing non è un'opzione di lusso per la ricerca e i ricercatori, ma è frutto di un servizio di controllo essenziale, laddove l'informazione senza filtri, senza "timbri virtuali di qualità", rimarrebbe sconosciuta e sarebbe inutilizzabile.

E' importante sottolineare che gli esperti che effettuano la valutazione dei lavori scientifici per le riviste sono gli stessi ricercatori che solitamente producono i lavori. E come gratuitamente rendono conto delle loro ricerche cedendo i propri lavori scientifici agli editori, altrettanto gratuitamente svolgono attività di peer-review.

Durante le passate decadi, da quando il numero di lavori scientifici ha subito l'incremento notevole che conosciamo, il refereeing dei lavori è divenuto quanto mai indispensabile appunto a causa dell'incremento di produzioni incerte dal punto di vista qualitativo, questo indipendentemente dal Web. Come afferma Renato Spigler³⁸, il problema è che il processo di peer-review di un *paper* richiede il coinvolgimento di almeno un esperto e del suo tempo e, per certi argomenti specialistici, vi sono davvero pochi esperti che possono dedicarsi a quel *paper*.

La situazione che si è creata, anche fuori dal Web è che vi sono "troppi lavori e troppo pochi buoni esperti" e questo rende difficile assegnare un giudizio appropriato ad un *paper* in un ragionevole lasso di tempo. Ciononostante, afferma il direttore esecutivo dell'American Mathematical Society (AMS)³⁹, qualcosa deve essere fatto: "meglio poco che niente".

Del resto, come osserva Paul Ginsparg⁴⁰, gran parte del tempo editoriale dovrebbe essere speso nella valutazione dei lavori rifiutati. E' impossibile evitare questo carico indispensabile,

³⁸ Renato Spigler, *Peer-Reviewing and Electronic Publishing*, "High Energy Physics Libraries Webzine", 6 (2002), <<http://library.cern.ch/HEPLW/6/papers/5/>>.

³⁹ John H. Ewing, *No Free Lunches: We Should Resist the Push to Rush Research Online*, "Chronicle of Higher Education", October 12, 2001, <<http://www.chronicle.com>>.

⁴⁰ Paul Ginsparg, *Creating a global knowledge network*, Invited Contribution for Conference held at UNESCO HQ, Paris, 19-23 February, 2001, Second Joint ICSU Press - UNESCO

paragonabile al "tempo impiegato dallo scultore nel rimuovere il materiale inessenziale" per dare forma all'oggetto nella sua interezza.

“Alcuni anni fa si diceva che, in letteratura matematica, ogni articolo specializzato era letto, in media, da non più di 1,5 lettori (e qualcuno maliziosamente sosteneva che questo numero includesse anche i revisori)”⁴¹.

In quest'ottica diviene subito chiaro che il solo vero costo per gli editori delle riviste scientifiche in merito al controllo di qualità è quello dell'organizzazione del processo di valutazione e non la sua effettiva esecuzione.

Va quindi sganciata, come sostiene Harnad, la fase dell'organizzazione del processo, di competenza dell'editore, da quella dell'effettiva esecuzione, che si riconduce ad un processo intellettuale altrettanto gratuitamente ceduto dai ricercatori delle università agli editori.

In ogni caso, è bene porre l'accento sull'importanza del processo QC/C, che comunque potrebbe anche essere attuato entro gli archivi aperti, anche se con meccanismi necessariamente diversi: certificare un lavoro dentro un Open Archive non necessariamente conduce ad un insieme omogeneo di lavori di qualità come può essere considerata una buona rivista scientifica.

Verso il cambiamento

In questa dimensione *Open*, dove gli *Open Archive* si costruiscono con programmi *Open Source* e si basano su protocolli *Open URL* entro *Open Digital Libraries* ove si respira aria di libertà, è difficile rientrare in un realtà in cui vengono imposte clausole contrattuali che limitano diritti fondamentali, come per esempio la *Regola di Ingelfinger*⁴².

La *Regola di Ingelfinger*, nata sotto l'egida di Franz Ingelfinger, direttore editoriale del "New England Journal of Medicine", adotta una politica editoriale che rifiuta di pubblicare o effettuare peer review su quei lavori che sono stati precedentemente pubblicati - o anche solo pubblicizzati - altrove (per esempio in convegni). Alcuni periodici biomedici hanno adottato questa regola, che serve agli editori per proteggere le loro fonti di guadagno, mentre altri, come ad esempio "Nature", l'hanno già eliminata perché, essendo in contraddizione con gli interessi della ricerca, nuoce all'immagine stessa della rivista.

Expert Conference on Electronic Publishing Science, during the Session "Responses from the scientific community", February 20, 2001, <<http://arxiv.org/blurp/pg01unesco.html>>.

⁴¹ Renato Spigler, cit.

⁴² Stevan Harnad, *Ingelfinger Over-Ruled: The Role of the Web in the Future of Refereed Medical Journal Publishing*. "The Lancet Perspectives" 256 (2000) 16, <<http://cogprints.ecs.soton.ac.uk/~harnad/Papers/Harnad/harnad00.lancet.htm>>.

La *Open Society* ha bisogno di poggiare su meccanismi *Open Access*⁴³, e per questa ragione sono fondamentali iniziative come quella di Soros o quella del World Health Organization (WHO)⁴⁴, la quale rende disponibili liberamente in formato elettronico oltre 1.000 periodici commerciali pubblicati dai maggiori editori contenenti i lavori di ricercatori provenienti da oltre 70 paesi in via di sviluppo. Anche editori commerciali come BioMed Central⁴⁵ stanno pubblicando periodici Open Access.

Cosa può fare la comunità scientifica per accelerare il cambiamento che comunque sarà ineluttabile? Tutti i ricercatori possono già da ora liberare la propria ricerca, come suggerisce Harnad, prendendo in mano la situazione e impossessandosi della propria ricerca.

I diritti di abbonamento, se eccessivi, sono ostacoli all'accesso e dunque all'impatto e, di conseguenza, a una libera ricerca. L'abbonamento è il mezzo per un editore di una rivista di rifarsi delle spese e di garantirsi un profitto equo, ma spesso non si tratta di un profitto equo.

E' tempo ormai che ogni ricercatore si ponga nell'ottica di fare in modo che certi meccanismi distorti vengano riequilibrati a favore della ricerca a tutela del suo stesso lavoro, ma anche degli interessi dell'Università in cui opera. Questo è già possibile con l'auto-archiviazione negli archivi elettronici.

Uno dei modelli più moderni e sofisticati, ma altrettanto facili da implementare e gestire è il server Eprints di Stevan Harnad⁴⁶, che presenta numerose funzionalità e potenzialità davvero sorprendenti e lascia ampie possibilità di configurazioni specifiche per esigenze personalizzabili: difatti l'interfaccia di staff e quella per l'utente sono agevoli e intuitive.

L'auto-archiviazione secondo questo modello Open Archive è rapida e facile, ed il sistema consente di caricare, a fianco della prima versione, eventuali aggiornamenti con procedure automatiche. E' possibile inoltre limitare la visibilità e l'accesso nel caso di documenti "protetti" a singoli gruppi, allo staff, ai soli utenti registrati.

Auto-archiviare la pre-pubblicazione è la prima misura indispensabile da effettuare, prima ancora di aver sottomesso l'articolo a una rivista, in quanto la proprietà intellettuale non appartiene che agli autori stessi, i quali non la tengono per sé ma la pongono a disposizione della comunità.

Il sistema gestisce qualsiasi formato di documento e offre funzionalità di browsing e ricerca assai sofisticate. Le versioni OA messe a disposizione dal gruppo di Harnad sono costruite

⁴³ Alexander M. Grimwade, *Open Societies Need Open Access*. "The Scientist" 16 (2002) 4, <http://www.the-scientist.com/yr2002/feb/comm_020218.html>.

⁴⁴ <www.healthinternet.org>.

⁴⁵ <www.biomedcentral.com>.

⁴⁶ <www.eprints.org>.

con programmi a codice sorgente aperto, per l'implementazione di e-server istituzionali, disciplinari o multidisciplinari, secondo le specifiche del protocollo Open Archive Initiative. Tali software sono liberi, rapidamente e facilmente installabili. Ecco come il termine *Open*, come per magia, apre le porte della conoscenza.

Implementare un e-server è una tappa importante, anche se non è sufficiente a creare degli archivi elettronici; in ogni caso è necessario che tutti i ricercatori, per quanto possibile, comincino ad auto-archiviare i propri articoli. Le biblioteche e i bibliotecari possono essere un punto di riferimento, un supporto organizzativo all'archiviazione stessa dei documenti, in stretta sinergia con gli informatici.

E' innegabile che ci troviamo di fronte ad una trasformazione che investe il concetto stesso di archivi aperti, il quale viene percepito come un mondo nuovo: da *pre-print* l'Open Archive si trasforma in *e-print*, vale a dire perde il prefisso "pre" per ricomprendere tutti i tipi di documenti. E questo è il vero Open Archive: un deposito di documenti appartenenti a una determinata comunità. Il ruolo dell'*EPrints* di Harnad è assai più ampio di quello inteso fino a poco tempo fa, quando ci si fermava al concetto di pre-print.

Anche in relazione alla proprietà intellettuale il pre-print è forse considerato "no man land" e pertanto meno compromettente da gestire rispetto ad altri documenti ceduti ad "altri territori". Nell'Open Archive di Harnad tutti i documenti trovano posto e tutti i formati sono ospitabili. Nella visione di Harnad il richiamo ai concetti di proprietà intellettuale è forte e ben radicato e per questa ragione il suo modello propone dei meccanismi considerati "sovversivi": combatte gli elementi negativi dell'applicazione di un copyright iniquo con le stesse armi del copyright, spesso basandosi appunto sui diritti di proprietà degli autori e sul loro consenso.

In Harnad abbiamo non solo una filosofia di fondo rielaborata attorno al concetto di opera, molto vicina ai nostri concetti biblioteconomici, ma esiste una forte componente che si basa sul senso di appartenenza ad una determinata comunità. L'atto del depositare sul server rappresenta uno sganciamento dai vecchi archetipi, ma al contempo è un atto di libertà che racchiude in sé tutti gli elementi di "sottomissione"⁴⁷ alla comunità stessa, intesa nella sua forza comunicativa di disseminatrice di conoscenza.

⁴⁷ Dal termine inglese "submission" che, entro un OpenArchive, ha significato di sottomissione del lavoro all'approvazione della comunità che afferisce a quell'archivio. La sottomissione del papers è effettuata dallo stesso autore che ne chiede il deposito sul server.