Open Access Publications

Giandomenico Sica
Editor

Open Access: Open Problems

Polimetrica
International Scientific Publisher

2006 Polimetrica ® S.a.s.
Corso Milano, 26
20052 Monza – Milano
Phone ++39. 039.2301829
Web site: www.polimetrica.com

ISBN 88-7699-053-4 Electronic Edition

Table of Contents

G. Sica (ed.) *Open Access: Open Problems*

Preface

Giandomenico Sica

Polimetrica is a company which aims to develop open access products. In 2005/2006 we've published the following open access works:

- Valore, Paolo, ed. (2006) Topics on General and Formal Ontology.
- Bachschmid, Nicolò and Pennacchi, Paolo, ed. (2006) Advances in vibration control and diagnostics.
- Gianini, Gabriele and Sillitti, Alberto, ed. (2006) Sharing ExperienceS on Agile Methodologies in Open Source Software development (SESAMOSS).
- Pitrelli, Nico and Sturloni, Giancarlo, ed. (2005) La stella nova. Atti del III Convegno Annuale sulla Comunicazione della Scienza.
- Mazzonetto, Marzia and Merzagora, Matteo and Tola, Elisabetta (2005) Science in Radio Broadcasting. The role of the radio in science communication.
- Rotolo, Giuseppe and Primiero, Giuseppe (2005) Dall'Artificiale al Vivente. Una storia naturale dei concetti.
- Cesareo, Vincenzo, ed. (2006) The Eleventh Italian Report on Migrations 2005.
- Baggio, Fabio and Zanfrini, Laura, ed. (2006) Migration Management and Ethics. Envisioning a Different Approach.
- Soro, Alessandro and Armano, Giuliano and Paddeu, Gavino, ed. (2006) Distributed Agent-Based Retrieval Tools. Proceedings of the 1st International Workshop.
- San Vicente, Félix, ed. (2006) Lessicografia bilingue e traduzione: metodi, strumenti e approcci attuali.
- Pitrelli , Nico and Sturloni, Giancarlo, (ed.) (2006) La scienza nella società del rischio. Atti del IV convegno nazionale e sulla comunicazione della scienza.
- AA.VV, (2005) Atti del 4° Congresso Nazionale dell'Associazione Italiana di Fisica Medica.
- Merlo, Maurizio (2006) La legge e la coscienza. Il problema della libertà nella filosofia politica di John Locke.
- Mazzolani, Federico M., ed. (2006) Seismic upgrading of RC buildings by advanced techniques - The ILVA-IDEM Research Project.

- Mazzolani, Federico M., ed. (2006) Innovative Steel Structures for Seismic Protection of Buildings - PRIN 2001.
- Seddone, Guido (2006) Condivisione ed impegno. Linguaggio, pratica e riconoscimento in Brandom, Hegel e Heidegger.

We've done this thanks to the kind support of different academic partners:

- Università Statale di Milano.
- Politecnico di Milano.
- Sissa di Trieste.
- Fondazione ISMU.
- CRS4.
- Università Statale di Bologna.
- Università Statale di Padova.
- Università Statale di Napoli.

In the next future we would like to improve this editorial policy, by extending it also to peer reviewed scholarly audio, video and software documentation.

In order to achieve this goal, it's important to understand which can be the possible forms of open access publishing trade.

For this reason, we've asked some of the most important researchers in this field of study to investigate this subject, by proposing their own experience and point of view.

The result is this book, which is not an arrival but a step in a long and articulated way directed to the definition of a valid open access editorial mechanism.

Concluding, I'd like to thanks all the people in Polimetrica who have worked and are still working in order to allow the development of this editorial vision and in particular Giovanni Sica, Polimetrica's Administrator and main inspirer of this policy, who each day finds the energies to have dialogues and discussions with an academic world which often sees open access as a beautiful word but not as a possible concrete reality and, most of all, as an important opportunity for the development of science.

Giandomenico Sica
Scientific Director
POLIMETRICA

Open Access al centro dei nuovi scenari di e-governance

Antonella De Robbio

Abstract. Il presente lavoro è focalizzato sull'Open Access o "Accesso Aperto". L'articolo puntualizza alcuni concetti fondamentali, definizione, scopi e obiettivi e ruolo dell'accesso aperto entro i modelli di comunicazione scientifica formale e informale. Con il termine Open Access si intende l'Accesso aperto alle produzioni intellettuali dei ricercatori e degli studiosi di tutto il mondo. L'Open Access è un movimento che incoraggia scienziati, ricercatori e studiosi a disseminare i propri lavori di ricerca rendendoli liberamente accessibili alle altre comunità di ricerca. Lo scopo dell'Open Access è rimuovere ogni barriera economica, legale o tecnica all'accesso dell'informazione scientifica, ciò al fine di garantire il progresso scientifico e tecnologico a beneficio di tutta la collettività. Una delle barriere è il copyright, che attualmente si basa su sistemi normativi che non agevolano lo scambio e la ricerca scientifica, bloccando l'accesso all'informazione di contenuto scientifico e chiudendola entro piattaforme editoriali a pagamento. Altro pericolo è dovuto all'attuale configurazione del mercato in cui si muove la ricerca scientifica oggi, dove vi è il rischio che grosse società o multinazionali acquisiscano il monopolio di prodotti ad alto rendimento commerciale a danno del progresso scientifico e tecnologico.

Il termine Open Access o Accesso Aperto, nasce nel contesto internazionale della ricerca, si sta sviluppando in Europa da alcuni anni e recentemente è ormai giunto anche in Italia.

G. Sica (ed.) *Open Access: Open Problems*

È un movimento internazionale che incoraggia scienziati, ricercatori e studiosi a disseminare i propri lavori di ricerca rendendoli liberamente accessibili secondo due modalità:

- depositando il proprio lavoro o ricerca scientifica in un archivio aperto attraverso un processo noto come self-archiving o auto-archiviazione;
- pubblicando il proprio lavoro o ricerca scientifica su periodici ad accesso aperto, ossia quei periodici che offrono gratuitamente e senza restrizioni l'accesso agli articoli, a seguito di regolare processo di validazione (referaggio) in termini di qualità.

Con Accesso aperto alla letteratura scientifica si intende l'accesso libero via Internet alle produzioni intellettuali dei ricercatori e degli studiosi di tutto il mondo. Esiste anche una definizione ufficiale di Pubblicazione ad Accesso Aperto, nota come Bethesda Statement on Open Access Publishing, abbracciata e condivisa anche da PLoS, Public Library of Science, da vari atenei e istituzioni britanniche e statunitensi, dallo statement del Wellcome Trust in supporto all'open access publishing e anche dall'IFLA, la Federazione Internazionale delle associazioni bibliotecarie.

Il mezzo digitale oggi ci offre la possibilità non solo di distribuire le produzioni intellettuali della ricerca, ma di disseminarle.

Il concetto di distribuzione, legato ad un mercato di editoria tradizionale a stampa, implica un prodotto che può essere distribuito a un numero discreto – inteso in senso matematico – di utenti. La distribuzione avviene comunque in differita rispetto ai tempi dell'effettiva produzione e confezione di un'idea: talvolta passano anni dal momento della produzione intellettuale di nuove idee o scoperte alla loro concreta pubblicazione a stampa in un periodico. Il mezzo digitale non solo diffonde l'informazione comunicandola a "enne" utenti al di là dello spazio e del tempo, ma la dissemina producendo una massimizzazione dell'impatto delle produzioni entro le comunità. Si attua quel fenomeno noto come "intelligenza collettiva" ben descritto da Pierre Levy, filosofo di cultura virtuale contemporanea che insegna al Dipartimento di Hypermedia all'Università di Paris VIII, a Saint Denis.

G. Sica (ed.) *Open Access: Open Problems*

Entro una cornice di intelligenza collettiva gli studiosi si scambiano idee entro il cyberspazio in un continuo colloquio di crescita collettiva continua. Ma questo accadeva anche prima del web. Un lavoro di Enrico Fermi sull'emissione di neutrini nel decadimento beta e sottoposto al periodico internazionale *Nature* verso il 1933, fu rifiutato dal comitato editoriale con la motivazione "di contenere speculazioni troppo remote per essere di reale interesse per il lettore". Ciononostante, le idee di Fermi circolarono su un doppio canale, pubblicate in italiano su riviste considerate minori in quanto "locali", mentre a livello internazionale il paper rifiutato da *Nature* (ma pare anche da altri giornali) circolò ovunque come report tecnico informale. Nonostante il suo circuito comunicativo fosse per lo più basato su letteratura informale, negli anni a seguire e fino agli esperimenti condotti da parte di Clyde Cowan e Fred Reines nel 1953 su reattore nucleare e successivamente confermati nel 1956, era ancora viva la polemica attorno alle scoperte precedenti sui neutrini di Enrico Fermi negli anni Trenta.

Inoltre quando parliamo di medium digitale parliamo di fluidità. Come ci ricorda Zygmunt Bauman, il sociologo polacco di *Modernità liquida*, la fluidità è lo stato dei liquidi e dei gas; un corpo fluido, a differenza di uno solido, può mutare continuamente forma se soggetto a forze o pressioni La fluidità, secondo Bauman, è la principale metafora dell'attuale fase dell'epoca moderna. Le informazioni digitali viaggiano, come i fluidi, ad estrema velocità; questo può essere un grande vantaggio per il progresso scientifico e tecnologico, basta saper cogliere le giuste opportunità che il mezzo digitale ci offre.

Possiamo affermare che l'*Open Access* non è solo un movimento, un insieme di iniziative internazionali con al centro gli scienziati e i bibliotecari coalizzati assieme, ma una strategia.

Gli attuali meccanismi che regolano i processi della comunicazione scientifica, ben noti ai ricercatori e agli studiosi, di fatto ostacolano l'impatto delle produzioni intellettuali di ricerca entro le comunità scientifiche.

L'Open Access combatte il paradosso della proprietà intellettuale nel circuito della comunicazione scientifica che ostacola i processi di crescita e sviluppo della scienza, tentando al contempo di arginare l'emorragia della spesa per la letteratura scientifica.

G. Sica (ed.) *Open Access: Open Problems*

Ogni anno vengono pubblicati circa due milioni di articoli in ventimila riviste, tenuti "prigionieri" entro riviste scientifiche a pagamento. Un periodico Open Access non chiede agli autori la cessione dei diritti economici o copyright.

L'Open Access riguarda la comunicazione scientifica e la disseminazione delle produzioni intellettuali di ricerca non l'editoria di varia o la musica o lo spettacolo. Inoltre si riferisce ai contenuti digitali e non alla carta o ai supporti analogici, pertanto eventuali timori o allarmismi provenienti dai settori dell'editoria di varia o dal mondo della distribuzione libraria sono del tutto infondati.

I papers pubblicati su riviste, a causa dei costi sempre più elevati e delle clausole di copyright sempre più restrittive, che ne impediscono anche la libera riproduzione, non sono disseminati come dovrebbero. La disseminazione dei papers è la funzione primaria ai fini di una fruizione e conseguente citazione da parte di altri papers, a tutto vantaggio del progresso scientifico e tecnologico collettivo. Qualsiasi limitazione o rallentamento dei processi della disseminazione influenza negativamente l'impatto sulla comunità, con pesanti ricadute cultuali, sociali ed economiche.

Le pubblicazioni di tipo tradizionale oggi sono invece una barriera all'accesso dei lavori scientifici. Per i settori scientifici la disponibilità immediata della letteratura prodotta è ormai una necessità non più prorogabile.

I preprint e i lavori intellettuali considerati tradizionalmente come "letteratura grigia", rapporti tecnici, relazioni ai convegni, atti di congressi, documenti progettuali, documenti pre e post pubblicazione, sono una fonte preziosa per lo sviluppo di qualsiasi settore disciplinare, vedi il citato esempio di Fermi sopra riportato. Molto spesso però questa tipologia di materiale non rientra per nulla nei canali della distribuzione a stampa e rimane così "nascosta" e priva di un effettivo impatto nella comunità dei parlanti".

Solitamente un pre-print evolve entro un articolo di periodico o entro un volume o entro un'altra forma di pubblicazione a stampa o elettronica. Tale evoluzione, soprattutto nelle pubblicazioni a stampa, comporta che il paper diviene difficilmente accessibile a causa dei prezzi sempre più alti degli abbonamenti alle riviste, senza contare i ritardi nei tempi di pubblicazione.

L'inaccessibilità ai contenuti intellettuali fa capo a due ostacoli.

G. Sica (ed.) *Open Access: Open Problems*

Il primo è relativo ai prezzi – sempre più alti - di abbonamento alle riviste, costi che influenzano in modo negativo il fattore di impatto entro la comunità scientifica e che aumentano il Digital Divide.

Il secondo è il blocco che concerne i diritti di proprietà intellettuale "degli autori" che finiscono gratuitamente nelle mani degli "editori", i quali, invocando la legge sul copyright, non consentono la duplicazione dei contenuti.

Per tali ragioni è nato il movimento internazionale Open Access il cui scopo è rimuovere ogni barriera economica, legale o tecnica all'accesso dell'informazione scientifica, ciò al fine di garantire il progresso scientifico e tecnologico a beneficio di tutta la collettività.

Come abbiamo detto all'inizio due sono quindi i canali dell'Open Access:

1. deposito negli Open Archives delle produzioni di ricerca;
2. pubblicazioni entro iniziative di editoria elettronica sostenibile.

A livello organizzativo gli OA si suddividono in:

- **Open Archive istituzionali**: l'archivio raccoglie, quale testimonianza della produzione intellettuale dell'ente, tutti i lavori di un particolare ente (università, ente di ricerca, dipartimento, …) o una parte selettiva dei lavori che l'ente ritiene di "conservare" nel deposito. In questo caso i materiali raccolti coinvolgono varie discipline.
- **Open Archive disciplinari**: l'archivio raccoglie i lavori in una determinata disciplina. Può anche trattarsi di un server di un ente che decide di aprire più archivi per discipline differenti. Molto spesso però si tratta di più soggetti (enti o anche soggetti individuali, dipende dall'organizzazione che si vuole adottare) che interagiscono nel deposito di materiale di una stessa disciplina o argomento specifico.

A proposito dei server disciplinari o subject-based, va sottolineato che essi si dividono in open archive:

- **a modello accentrato** (del tipo arXiv): unico grande server a carattere nazionale o internazionale
- **a modello distribuito** (del tipo la rete RePEC Research Papers in Economics[1] per gli economisti): piccoli server locali sparsi

Spesso i piccoli server non sono integrati entro una rete a modello distribuito - come avviene nel caso del modello *RePEC* - e per questa ragione molti piccoli server sono stati, nel corso degli ultimi cinque anni, assorbiti da server più grandi (per esempio da *arXiv*).

La *Open Archive Initiative OAI* si occupa di approntare standard e protocolli per l'interoperabilità tra archivi, per questo si parla di compatibilità OAI.

Nella terminologia *Open Archive Initiative (OAI)*[2]*,* gli e-print server sono detti "data provider" o "fornitori di dati", laddove per dati si intendono sia i "testi pieni" (full-text) sia i "dati sui dati" ovvero i metadati. L'architettura OAI prevede, a fianco dei data provider, anche dei "service provider", ovvero "fornitori di servizio", i quali sono preposti alla creazione di servizi a valore aggiunto sui data provider: accesso esteso, creazione di indici di più data provider, meta-motori per la ricerca multipla in più archivi contemporaneamente...

Nel mondo vi sono centinaia di eprints server OA; sono sorti attorno agli anni novanta sulla scia dell'antenato *ArXiv*[3] di LANL, messo in piedi da Paul Ginsparg, per la fisica. *ArXiv* ora comprende anche papers di matematica e informatica. Sebbene per la fisica esso sia l'OA per eccellenza, ve ne sono molti altri, per esempio quelli del CERN messi a disposizione dal *CDS Server*, *SPIRES/SLAC* di Stanford, il servizio giapponese *Kiss*[4] *KEK Informazion Service System*, e così via e i server della SISSA di Trieste.

Per la matematica, oltre all'ospitalità offerta dai "cugini" di *arXiv* a Los Alamos, e da cui è nato *Front for the Mathematics,* con oltre

[1] <http://repec.org/>.

[2] **Antonella De Robbio** *Open Archives Initiative (OAI) in Europa: Workshop al CERN di Ginevra*. In "Biblioteche Oggi". Maggio 2001, pp. 66-69.

[3] <http://arXiv.org/>.

[4] <http://www-lib.kek.jp/KISS/kiss_prepri.html>.

10.000 lavori[5], la mappa geografica si presenta frammentata e costellata da numerose esperienze, in particolare composta da server di piccole dimensioni di tipo disciplinare[6].

I matematici hanno sviluppato progetti per la creazione di *repositories* per il deposito dei loro lavori[7]; in ambito europeo esiste il progetto *MPRESS/MathNet.prEPrints server*[8], coordinato dai matematici tedeschi Judith Plümer, Roland Schwänzl.

MPRESS 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.

Per la medicina esistono esperienze nuove, iniziate a partire dalla fine del 2000, in connessione con prestigiose banche dati, come nel caso dell'integrazione tra la banca dati *PubMed (Medline)* e il repository *PubMedCentral*, tipo di archivio centralizzato che funziona attraverso accordi con editori. *NetPrint*, per il settore della medicina clinica, è molto più vicino ad un vero open archive, in quanto si basa sull'auto-archiviazione da parte degli autori.

Esistono anche Open Archive per la chimica, la biologia e per altre discipline, anche se, per quelle umanistiche c'è ancora molta strada da percorrere.

Cogprint di Stevan Harnad è l'OA per le scienze cognitive, piscologia, neuroscienze e linguistica dove i ricercatori in tre anni hanno depositato oltre mille articoli a testo pieno, entro un archivio che comprende oltre 6500 metadati, recuperati da altri archivi di bibliografie in BibTex[9].

Accanto agli e-print server disciplinari, stanno sorgendo, a livello internazionale, e-print server istituzionali, politicamente più forti, rispetto alle aggregazioni libere messe in atto, in questi anni, da

[5] Front End for the Mathematics ArXiv, http://front.math.ucdavis.edu/>.

[6] Vedi il repertorio *Mathematics and Physics Preprint and e-Print Servers* by Antonella De Robbio <http://library.cern.ch/derobbio/mathres/preprint.html>.

[7] <http://www.ams.org/global-preprints/umbrella-server.html>.

[8] MPRESS - MathNet.preprints, <http://MathNet.preprints.org/>.

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

comunità disciplinari; essi infatti richiedono un'organizzazione stabile entro un quadro politico di responsabilità ben definita.

Uno dei due canali dell'Open Access riguarda infatti la costruzione di archivi aperti (Open Archives) entro le università e gli enti di ricerca, utili alla raccolta, al deposito, e conseguente disseminazione del materiale prodotto dalla ricerca. Gli archivi di e-prints sono server che consentono agli autori di rendere i propri lavori liberamente disponibili alla comunità internazionale scientifica, disseminandoli su scala mondiale, cosa impossibile per un lavoro su carta. La procedura definita come "self-archiving", prevede l'invio dell'articolo da parte dell'autore ad una rivista peer-reviewed (tradizionale o a modello open access) contestualmente al deposito nell'archivio. Attraverso il self-archiving o auto-deposito, gli autori alimentano gli archivi e i bibliotecari controllano la correttezza dei metadati, mentre la qualità dei dati o contenuti è garantita dalla sottomissione degli articoli alle riviste. Questa pratica è consentita dalla maggioranza degli editori, o è comunque contrattabile. Grazie alla tecnologia web e alla disponibilità di adeguati software (solitamente open source) per la gestione degli e-prints, è possibile implementare un Open Archive di documenti elettronici rendendo i documenti prodotti dagli studiosi liberamente accessibili e a disposizione della comunità. In questi luoghi è possibile depositare i papers scientifici, i lavori intellettuali, e tutti i documenti che si ritiene utile mettere a disposizione delle comunità scientifiche.

In questo modo i server istituzionali, di fatto, si contrappongono all'attuale modello distorto del circuito della comunicazione scientifica, imponendosi come luoghi per il semplice deposito dei materiali.

Il secondo canale dell'Open Access riguarda invece la creazione, gestione di riviste "aperte". Solitamente in una rivista open access gli autori ritengono il copyright e quindi l'autore può fruire liberamente della propria produzione intellettuale. L'autore la può depositare immediatamente in un archivio aperto, rendendola immediatamente disponibile anche prima del processo di referaggio, per una libera e aperta discussione con tutta la comunità. Esiste una banca dati internazionale, gestita dall'Università di Lund, che censisce tutte le riviste open access di qualità in ogni settore disciplinare.

G. Sica (ed.) *Open Access: Open Problems*

Ad oggi DOAJ Directory Open Access Journal[10] enumera quasi 2000 e-jornal open access.

Esistono strumenti tecnologici che consentono di approntare piattaforme e server "aperti" compatibili con i protocolli e gli standard internazionali di interoperabilità.

Questi modelli poggiando su software libero sfruttano la filosofia del 'copyleft', giuridicamente basato sul copyright, che consente una piena e totale libertà di utilizzo da parte degli utenti, nel rispetto delle tutele poste degli autori.

Tutta la ricerca dovrebbe passare attraverso i canali dell'accesso libero per il semplice motivo che nessuno paga gli autori per i loro lavori scientifici. Le produzioni intellettuali dei ricercatori dovrebbero essere, 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". È ormai indiscusso che per esserci impatto è necessaria un'ampia disseminazione, in altri termini i lavori dei ricercatori devono essere letti, citati e utilizzati da altri ricercatori, solo così raggiungono l'impatto utile ad uno sviluppo collettivo, quell'impatto che consente di creare nuove ricerche, di effettuare nuove scoperte sulla base di un lavoro altrui letto, assimilato, metabolizzato.

Il copyright è degli autori e gli autori dovrebbero fare attenzione a non cedere diritti fondamentali a terzi, i quali chiudono gli accessi entro piattaforme a pagamento. Abbiamo riviste scientifiche che costano anche 22.000 dollari di abbonamento annuo e i lavori pubblicati in tali "fortezze chiuse" sono lavori regalati dagli autori agli editori, attraverso un meccanismo di give-away che priva del tutto autori e istituzioni scientifiche di ogni diritto legato a quell'articolo. Non è più possibile riusare tali lavori per eventuali pubblicazioni in saggi o per convegni o anche per fini didattici. Attualmente abbiamo leggi capestro che mettono legacci da tutte le parti e che sono ritagliate attorno a modelli che nulla hanno a che fare con la ricerca scientifica, per esempio musica e spettacolo.

Il mercato dei supporti si aggrega attorno a lobbies potenti in grado di influenzare i governi nella scrittura di leggi e decreti a loro

[10] DOAJ Directory of Open Access Journals http://www.doaj.org/.

favore piuttosto che a favore della collettività e purtroppo questo accade sia a livello italiano, sia a livello europeo (basti vedere le direttive UE sempre più restrittive), sia a livello internazionale, dove ormai l'attività della WIPO, World Intellectual Property Organization, e di OMPI, Organizzazione Mondiale Proprietà Intellettuale, sta facendo gli interessi delle grosse potenze (USA e Europa). Il termine pirateria ne è l'esempio più lampante. Significa fare una copia non autorizzata di un'opera, di un lavoro. Se trasponiamo questa libera azione nel contesto della comunicazione scientifica, laddove il fare una copia significa disseminazione in termini di impatto - presupposto fondamentale ad una crescita produttiva e sociale che passa necessariamente attraverso una crescita culturale - ci accorgiamo subito di come il termine pirateria, entro un quadro di comunicazione scientifica, sia del tutto fuori luogo. Quindi da una parte abbiamo la normativa sul diritto d'autore che rende tutto inaccessibile in quanto "protetto", dall'altra abbiamo il pubblico dominio, ovvero la non protezione totale. Nel pubblico dominio l'opera non ha alcuna tutela e questo non va bene per i lavori di ricerca, che devono comunque trovare la loro giusta dimensione anche in termini di diritto morale di autore. Di mezzo si collocano le licenze, nella sfera del copyleft, mutuato dal software libero. Un autore decide in modo del tutto autonomo cosa l'utente può o non può fare con il lavoro da lui creato e messo in rete. L'autore attaccando una licenza, per esempio del tipo Creative Commons, pur mantenendo in pieno il suo diritto morale come autore, decide di liberare il proprio lavoro in merito a riproduzione e distribuzione, in barba a termini come "pirateria", mal adattabili al contesto della ricerca. Può decidere anche in merito ad altri diritti, come quello di rielaborazione o di uso commerciale. La licenza ha tre volti, uno per l'utente normale, che è subito avvertito in modo semplice, chiaro e sintetico dei termini della licenza, uno ad uso legale e il terzo si connota come metadato standard, incarnato nella risorsa stessa e visibile ai motori di ricerca.

Certo è che se l'autore ha ceduto i diritti ad un periodico non Open Access difficilmente potrà attaccare una licenza di questo tipo al proprio lavoro. Gli archivi aperti si stanno muovendo verso l'adozione di licenze da proporre agli autori direttamente nella fase di deposito.

G. Sica (ed.) *Open Access: Open Problems*

Entro i nuovi modelli di mercato, la distinzione tra comunicazione formale e informale diverrà sempre più effimera.

La trasformazione della lineare e familiare catena dell'informazione scientifica, in una rete interattiva per la comunicazione scientifica, è la risposta, come ci dicono Roosendaaal e Geurts nel loro fondamentale articolo che risale al 1997[11], ai cambiamenti concomitanti nei campi della ricerca scientifica e della didattica.

Questo sviluppo necessita di nuove vie di gestione della conoscenza, tra cui la revisione delle quattro funzioni che caratterizzano il circuito della comunicazione scientifica: registrazione, awareness, certificazione, archiviazione, per giungere alla quinta funzione che concerne il rewarding.

Secondo Roosendaal e Geurts, le quattro funzioni "principe" che intervengono nei processi di comunicazione scientifica, sono quattro forze in movimento che consentono analisi strategiche entro il mercato. La loro interazione reciproca gioca un ruolo fondamentale nel ridisegnare nuove strutture di comunicazione e di pubblicazione.

Il filosofo francese Bourdieu[12] ci dice che l'universo della scienza è oggi minacciato da una terribile regressione a causa di un'autonomia che si sta indebolendo a seguito di meccanismi sociali, come la logica delle concorrenze di mercato, che rischia di mettere la scienza al servizio di fini imposti dall'esterno.

La WIPO stessa deve cambiare. Recentemente è stata stilata da parte di alcuni movimenti del contesto open un documento noto come la Dichiarazione di Ginevra sul Futuro dell'Organizzazione Mondiale per la Proprietà Intellettuale. Come organizzazione intergovernativa la WIPO si è messa nella direzione di creare ed espandere privilegi monopolistici, spesso senza badare alle conseguenze sociali ed economiche che una eccessiva protezione della proprietà intellettuale comporta. L'espansione continua di questi privilegi e dei loro meccanismi di applicazione ha causato gravi costi sociali ed economici e ha ostacolato e minacciato altri importanti sistemi per la creatività e l'innovazione. Esistono innovazioni

[11] **Hans E. Roosendaal**, **Peter A. Th. M. Geurts** "Forces and functions in scientific communication: an analysis of their interplay". <http://www.physik.uni-oldenburg.de/conferences/CRISP97/roosendaal.html>.

[12] Pierre Bourdieu, Il mestere di Scienziato

incredibilmente promettenti nel campo delle tecnologie dell'informazione, della medicina e di altri settori essenziali, così come all'interno dei movimenti sociali e dei modelli economici, per promuovere gli scambi e trasferire le conoscenze. È necessaria una moratoria sui nuovi accordi e sull'armonizzazione degli standard che espandono e rafforzano i monopoli e restringono ulteriormente l'accesso alla conoscenza. Per generazioni la WIPO ha risposto in primo luogo alle settoriali preoccupazioni di potenti editori, industrie farmaceutiche e altri interessi commerciali. Le ricerche pubblicate sui periodici scientifici e sulle basi di dati, le informazioni sul genoma umano, le ricerche sui farmaci per curare malattie endemiche come AIDS o le ricerche per sconfiggere per esempio le malattie autoimmuni devono essere un bene comune, un "common", per usare un termine anglossassone. L'istruzione a distanza deve essere un mezzo per trasferire le conoscenze derivate dai processi di ricerca entro modelli aperti di didattica distribuita. Il controllo della conoscenza, della cultura, della tecnologia, persino delle risorse biologiche da parte di pochi a scopo di lucro, danneggia uno sviluppo collettivo che rispetti in modo democratico le diversità che costituiscono la nostra ricchezza.

Va sottolineato che durante la *Inter-Sessional Intergovernmental Meeting (IIM)* della WIPO tenutasi a Ginevra il 20-22 luglio 2005, IFLA e e-IFL unite a numerose ONG non riuscirono a raggiungere un accordo sui punti fondamentali che riguardano le biblioteche, la ricerca e l'insegnamento a causa di resistenze da parte degli Stati Uniti e del Giappone. Componente chiave per la proposta di uno sviluppo di un'Agenda WIPO "sostenibile" è una chiamata per un trattato sull'Accesso alla Conoscenza, in sigla *A2K, Treaty on Access to Knowledge*. Un A2K sarebbe fondamentale non solo per le biblioteche, ma per la didattica e la ricerca in quanto a causa di interessi economici e di forti lobby di mercato che detengono diritti su opere di interesse pubblico, è sempre più difficile reperire l'informazione, usarla e creare così nuova conoscenza. Tale capacità è essenziale per lo sviluppo di nuova conoscenza e risiede – a livello normativo – su ciò che viene definito "eccezioni e limitazioni al copyright". Nell'ultima decade i trattati internazionali, come le direttive sopranazionali dell'Unione Europea, le varie legislazioni nazionali e i termini di alcuni Trattati sul libero commercio, noti

G. Sica (ed.) *Open Access: Open Problems*

come FTA *Free Trade Agreements*, hanno creato una tendenza verso la monopolizzazione e privatizzazione dell'informazione attraverso un'erosione sempre più ampia delle eccezioni e limitazioni ai diritti, ciò in particolare entro l'ambiente digitale.

Un accesso equo all'informazione per tutti è base imprescindibile per il consolidamento dell'educazione e per stimolare l'innovazione. È quindi necessario procedere con un trattato apposito che reindirizzi il corretto bilanciamento e stabilizzi un framework internazionale che sancisca le norme dalle quali il copyright protegga i diritti degli utenti tanto quanto il mantenimento di adeguate protezioni per i detentori dei diritti.

Ai recenti alla conferenza internazionale dell'IFLA WLIC World Library and Information Congress tenutosi a Oslo, dal 14 al 18 agosto scorso, si sono tenute due sessioni della CML *Committee on Copyright and other Legal Matters* che hanno messo in luce i gravi e urgenti problemi a cui siamo costretti a far fronte negli ultimi anni in relazione alla gestione dei diritti nelle attività di biblioteca, nella ricerca e nella didattica, problemi a cui la WIPO non ha posto la dovuta attenzione.

In tale sede molto si è discusso sulle clausole contenute negli accordi multilaterali sugli investimenti (MAI) (Multilateral Agreement on Investment) o negli accordi bilaterali per il libero commercio (FTAs) (Free Trade Agreements) che si discostano dagli obblighi previsti dagli accordi TRIPs (Aspetti legati al commercio dei diritti di proprietà intellettuale) del WTO, l'Organizzazione mondiale del commercio (OMC), questo per assicurare a Stati Uniti e Giappone un accesso e un controllo sempre maggiore alle loro imprese nei paesi in via di sviluppo. Gli interessi economici sono enormi, anche in relazione alle questioni di brevettabilità di beni considerati patrimonio della cultura e tradizione locale.

Va ribadito che la questione dei brevetti rientra in un ambito differente rispetto a quello del copyright. La proprietà intellettuale si divide in sue branche: una riguarda i marchi e i brevetti, la cosiddetta proprietà intellettuale di tipo industriale, la seconda riguarda la proprietà intellettuale artistica e letteraria, dove abbiamo anche la ricerca scientifica laddove non comporti innovazioni di ricerca e sviluppo tali da essere ricompressa entro accordi di segreto o brevetto industriale.

Ormai i confini tra questi due settori sono sempre più sfumati per certe discipline.

Il nuovo accordo noto come TRIPs-PLUS ne è l'esempio più lampante, esso si estende ben oltre le normative previste dall'accordo TRIPs. Ciò mina sensibilmente la flessibilità contenuta nel TRIPs riaffermata, oltretutto, dalla Dichiarazione di Doha in materia di accesso ai farmaci. Il recepimento di tali accordi da parte dei governi di almeno venti paesi africani (ora sotto pressione al fine di una loro sottoscrizione al TRIPs-PLUS) minacciano seriamente l'accesso ai farmaci sia in Africa o anche in India, ma non solo. È noto che il Sudafrica è il paese che vede un'altra percentuale di malati di Aids, i farmaci salvavita costano l'equivalente del salario mensile di un operaio. Il Sudafrica avrebbe tutte le infrastrutture per la produzione di farmaci a costi decisamente accessibili. Vi è stata una denuncia da parte delle multinazionali dopo che attivisti statunitensi, europei, asiatici, latinoamericani, africani e organizzazioni non governative avevano promosso una campagna di boicottaggio su tali farmaci, ma il governo non ha avviato un'autonoma produzione dei farmaci, bensì ha contrattato il prezzo di vendita con le multinazionali. Le regole sbilanciate sulla proprietà intellettuale dell'"Uruguay round" dei negoziati sul commercio, dettate dalle industrie farmaceutiche proibiscono ai vari paesi in via di sviluppo di produrre farmaci generici, rendendo inaccessibili in questi paesi molti medicinali di importanza cruciale, compromettendo il rispetto del diritto alla vita e alla salute.

La proprietà intellettuale, sia essa di ambito brevettuale o ristretto al solo copyright è quindi più che mai al centro di processi economici dove interessi di governi più forti impongono ai paesi più deboli accordi o trattati che danneggiano fortemente lo sviluppo economico, sociale e culturale di questi paesi.

Bourdieu ci parla di "deliri postmoderni" in cui le pressioni dell'economia si fanno ogni giorno più forti soprattutto in certi settori dove le produzioni intellettuali sono altamente redditizie. E anche qui parliamo sempre di proprietà intellettuale la quale è divisa in due grandi aree di intervento, o meglio in due diritti assoluti, l'uno che ricade entro il diritto d'autore o proprietà intellettuale artistica e letteraria (dove si trovano le pubblicazioni scientifiche), l'altro comprende i marchi e brevetti e ricade entro la proprietà intellettuale

G. Sica (ed.) *Open Access: Open Problems*

industriale. Il pericolo, sottolinea Bourdieu sta nella linea di frontiera tra la ricerca di base e la ricerca applicata, dai confini sempre più sfumati, nel momento in cui vi è il rischio "che gruppi di ricerca cadano sotto il controllo di grandi società industriali, attente ad acquisire, attraverso i brevetti, il monopolio di prodotti ad alto rendimento commerciale".

Occorre considerare che questo non è argomento che interessa solo i paesi in via di sviluppo, ma anche i paesi sviluppati, in quanto la conoscenza è un diritto universale, e la parità nell'accesso è un sostegno indispensabile per qualsiasi società democratica e per ogni società ed economia inclusiva.

Antonella De Robbio
Responsabile Biblioteca Digitale e Referente copyright
CAB Centro di Ateneo per le Biblioteche
Università degli Studi di Padova

A concrete step for building public electronic archives of reviewed papers

Takashi Kunisawa

The recent progresses of computer science and communications technologies have been changing both the practices and products of science. Internet-search for a key paper, or method is now the everyday practice of science and technology. A large amount of data is exchanged rapidly and worldwide among researchers in many fields, notably genomics, neuroscience, and geophysics. There is a growing interdependence among traditional scientific disciplines; data collected in one scientific discipline are likely to be used in other disciplines. Thus, access to and sharing of scientific data and information are even more important for the conduct and advancement of science.

Recognizing these changes in scientific practices and products, the Committee on Data for Science and Technology (CODATA, http://www.codata.org) has making efforts to establish the principle of full and open access to data for scientific and educational purposes. A resolution made in 1988 by its mother organization, the International Council for Science (ICSU), reads "Recommends all ICSU members to support the fundamental principle of full and open exchange of data and information for scientific purposes" [1]. This policy of open access is followed and expanded in the Berlin Declaration [2], in which open access is defined as a comprehensive source of human knowledge and cultural heritage and its contribution includes not only original scientific research results and raw data but also metadata, source materials, digital representations of pictorial

G. Sica (ed.) *Open Access: Open Problems*

and graphical materials and scholarly multimedia material. Furthermore, the necessity for the open access principle was emphasized in the recent Tunis World Summit on the Information Society (WSIS); its final document reads, "*by supporting educational, scientific, and cultural institutions*, including libraries, archives and museums, in their role of developing, providing equitable, open and affordable access to, and preserving diverse and varied content, including in digital form, to support informal and formal education, research and innovation".

A different line of policy consideration has extensively been developed by an OECD Follow-up group; they focus on research data obtained using public funds and recommend to ensure that both researchers and the public receive optimum returns on the public investments in research and to build on the value chain of investments in research and its data resource [3]. Coordinated efforts at national and international levels are needed to broaden access to data from publicly funded research and contribute to the advancement of scientific research and innovation. To this effect, Ministers adopted a declaration entrusting the OECD to work towards commonly agreed Principles and Guidelines on Access to Research Data from Public Funding [4].

The access to scientific publications generally poses less serious problems than availability of research data. PubMed Central (PMC) was created by the National Institute of Health (NIH) as a public repository of publications, which provides free and convenient access, linked with the popular literature database, PubMed. A lot of journals, which are published from academic societies, including the Proceedings of National Academy of Sciences, Nucleic Acids Research, and Journal of Biological Chemistry, have already deposited their content with PMC, and research articles published in these journals can be freely accessed, in general, after six months of publication. Further, a considerable number of open access journals are published from BioMed Central (http://www.biomedcentral.com/) and from the Public Library of Science (PloS, http://www.plos.org/). A unique activity is recognized in Japan. J-STAGE (http://www.jstage.jst.go.jp/browse/), which is fully supported by the Japanese government, provides academic societies in Japan with software, hardware/facilities, network, security and daily maintenance

G. Sica (ed.) *Open Access: Open Problems*

necessary for electronic publication without requiring any charges. Nearly 300 journals enjoy this support and are found on the J-STAGE server, 80% of which are freely accessed. J-STAGE automatically provides mutual links to the major databases, PubMed, ChemPort and CrossRef, enhancing its usability. The J-STAGE project is thus very useful and helpful to implement the open access practice in Japan. Similar projects are highly desirable in other countries, in particular in developing countries.

An important step toward a public repository of reviewed papers was taken by NIH. The NIH policy on enhancing public access to archived publications resulting from NIH-funded research requests and strongly encourages all investigators to make their NIH-funded peer-reviewed, author's final manuscript available to other researchers and the public through PMC immediately after the final date of journal publication [5]. For quality assurance publication of research findings and results should be subject to rigorous scrutiny through effective mechanisms of peer review. The open access archives PMC will gain more value if research papers based on public funds other than NIH are collected. If such a means of enhancing public access is taken in Europe, Asia and other regions, then such archives will become more comprehensive and gain more value. Furthermore, collaboration and integration among the regional archives will allow efficient indexing, searching and linking, by which their usefulness dramatically increases [6]. The outputs from current and future research should be preserved and remain accessible not only for the next few years but for future generations. International collaborations are prerequisite for the stable long-term preservation of research outputs. Moreover, it is efficient to develop on an international frame software systems necessary for electronic publication. Such international collaborations remind us of the triangle among GenBank in the U.S., EMBL in Europe and DDBJ in Japan, which are most successful and popular open access repositories of nucleic acid sequence data. There are strong pleas from both the commercial sector and sympathetic governments to strengthen intellectual property rights, and within each country the positive value of the open access principle tends to be regarded as secondary relative to short-term commercial interests [7]. Therefore, international organizations including ICSU and CODATA should

play a leading role in creating such international collaborations for open access. Thus, I would like to urge all scientists to take appropriate actions to their government and leading scientists involved in international organizations to provide opportunities to implement the open access principle of reviewed papers that are results of publicly-funded research. The realization of integrated international repositories of publicly-funded research outputs must open the door to implement the principle of open access to research data.

Bibliography

[1] Available from http://www.codata.org/resources/databases/data_access/policies.html, in which policy statements of a number of organizations on data issues are collected.

[2] http://www.zim.mpg.de/openaccess-berlin/berlindeclaration.html.

[3] Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., Laaksonen, L., Moorman, D., Uhlir, P., Wouters, P. [2004] Promoting Access to Public Research Data for Scientific, Economic, and Social Development. *Data Science Journal* 3: 135-152.

[4] Available from http://www.oecd.org/document/15/0,2340,en_2649_201185_25998799_1_1_1_1,00.html.

[5] NIH Public Access, available from http://publicaccess.nih.gov.

[6] Roberts, R.J., Varmus, H.E., Ashburner M., Brown, P.O., Eisen, M.B., Khosla, C., Kirschner, M., Nusse, R., Scott M., Wold, B. [2001] Building a "GenBank" of the published literature. *Science* 291: 2318-2319.

[7] Lubchenco, J., Iwata, S. [2003] Science and the Information Society. *Science* 301: 1443.

Takashi Kunisawa
Department of Applied Biological Sciences
Science University of Tokyo
Noda 278-8510, Japan

G. Sica (ed.) *Open Access: Open Problems*

Open Access: national policy initiatives as an alternative to personal commitment

Derek Law

Background

The Open Access debate has been running for well over a decade. Ten years ago at a major conference in Paris sponsored by UNESCO and ICSU, Joshua Lederberg, the eminent scientist and Nobel prize-winner talked of the impact of technology and said:

> *'Now what are some of the foreseeable consequences? I really have nothing to ask of the print publishers or of the "for profit" electronic purveyors. Unless they are very selective - and they sometimes will be - about their value added, they will fall of their own weight as scientists become empowered to manage their own communications without the benefit of intermediaries.'* [5]

A decade later we should be clear that, with the honourable exception of ArXiv in physics, this simply has not happened in mainstream science.

Throughout the intervening years tireless proselytising by a host of John the Baptist like figures from Paul Ginsparg to Stevan Harnad and institutionally through SPARC has been unceasing, has won many battles, has nailed declarations to the doors of the publishing establishment from Budapest to Berlin, has eroded the edifice of traditional scholarly communication, has moved the debate from the fringes of discourse to the mainstream, has probably won the argument, but so far has not won the war. A recent survey [12] has shown how far repositories have spread in some thirteen countries. It

G. Sica (ed.) *Open Access: Open Problems*

also shows a very complex patchwork of data types, software platforms and a typically very low level of deposit.

At the same time open access journals have grown in number. In December 2005, the Directory of Open Access Journals (DOAJ) lists almost 1900 open access journals.[2] But open access is still a long way from being at the heart of scholarly communication and is ranged against large commercial forces in the STM publishing area. Swan's recent major study [9] shows that self-archiving, open access and institutional repositories are now widely understood by academics. Her survey results showed that:

- 39% of respondents have self-archived "in one form or another"
- 2% have published in an OA journal
- 69% would deposit willingly, if mandated to do so by their employer or funder

However, these figures conceal a large number of worries, although admittedly the worries rest on largely anecdotal evidence. Firstly it is worrying that while 39% of respondents have self-archived "in one form or another" a trawl round any institutional website for personal archives might suggest that a significant proportion of this traffic rests on non-OAI compliant and unharvestable web-pages. Secondly, any prolonged exposure to the relevant mailing lists demonstrates a continuing and worrying inability of many participants to distinguish between Open Access Journals and Institutional Repositories. Thirdly, there is clear worry and/or confusion amongst researchers over copyright, peer review and citation counting. It is as easy to interpret the fact that 69% would deposit willingly if mandated as an abdication of responsibility as an embracing of repositories.

In summary significant progress has been made in developing understanding and ambition but self-archiving remains a minority activity. Harnad estimates that 15% of the journal literature is placed in institutional repositories. And while he remains unswerving in his goal, it is worth remembering that the journal literature is itself a subset of peer-reviewed academic published outputs. The problem of

bringing about true cultural and organisational change remains a major one.

In order to address this issue a quite different approach is being explored in some countries. To follow a military analogy it is perhaps best seen as a second front than an alternative plan of attack. It also seems possible that this approach is particularly suited to small countries with limited indigenous publishing industries. Be that as it may, the problem of embedding cultural change in the scientific community may be as readily tackled at government level as at the personal scientist level.

1. Cultural and Organisational Change through Institutional Initiatives

One step above the ambition to influence personal culture has been the move to change practice through the intervention of the funding agencies. This has been led from the biomedical area where initiatives such as that by the Wellcome Trust to mandate open access have been widely welcomed but have been seen as intensely political acts. In the United Kingdom this has been followed by the equally politicised attempt by the major funding agencies in Research Councils UK (RCUK) to mandate deposit. This has elicited a fierce backlash from the publishing industry. The draft policy has been significantly delayed and has been treated with a posture somewhere between scorn and indifference by the relevant government minister. It is popularly supposed that a major lobbying effort by the large publishers is hampering progress on acceptance of the RCUK policy. Even the Royal Society, which has a substantial publishing arm, has issued an attack on the RCUK policy, which appears to be driven by its publishing needs rather then an examination of the future of scholarly communication. [8]

While the position in the UK is described here, it is by no means exceptional. Moves by major grant awarding bodies in countries with major publishing industries lead at best to major battles with the publishing houses and at worst to misguided government interference on the grounds that a wealth generating industry is being threatened. The debate quickly degenerates into a battle in which the status quo is defended rather than the future defined.

G. Sica (ed.) *Open Access: Open Problems*

2. Cultural and Organisational Change by alignment with government policy

Thus far the debate on open access has tended to lie within very large countries. It has been suggested however [6] that the information experience of countries varies according to size and geography. It is then worth exploring whether the problem of embedding cultural change can be tackled in a different way in smaller countries. Clearly countries are at very different levels in their understanding and practice of the issues. Perhaps the first stage is when a small country decides to adopt a national information strategy in order to achieve government goals. This usually involves some combination of preserving threatened cultural values and/or an aspiration to align the country in some way with moves towards a knowledge society. An excellent example of this might be New Zealand.

New Zealand

A draft Digital Strategy was released in June 2004 for public feedback and discussion. It is intended to be a five year plan and has budgetary support and proper monitoring and evaluation components and links to longer term goals.

The Digital Strategy intends to set New Zealand's direction for the next five years. It sets out key actions over the next few years where budgets have already been committed. It puts in place a structure against which to evaluate our progress and will ensure we meet our longer term goals. The Digital Strategy is closely linked to other government priorities, such as the Growth and Innovation Framework and the Sustainable Development plan. The website for the plan [7] claims to have "… consulted extensively with businesses and industry groups, community and voluntary groups, health professionals and educators, researchers, and individuals. We received nearly 200 written submissions…." Clearly based on UK experience of five years earlier it stresses the importance of content**,** connection**,** and confidence, and the need to develop all of them at the same rate. A substantial emphasis of the programme is the preservation of Maori culture. When the final strategy was launched

in November 2005, a separate related event looked at institutional repositories and celebrated the launch of the first such repository in the country.

Australia

This may be contrasted with a programme of development in Australia, again a small country (in population terms) with a small publishing industry. Australian universities and in particular their libraries have been quick to see the merits of institutional repositories and have made steady progress since 2002 when the "Repository Agenda" was established, with several separate repository initiatives, including an e-prints collection and an archive of Asian material. Within a year DSpace had become a de facto standard and had emerged as an institutional framework for repositories but still on a developmental basis. By bringing together the repository work with the Australian National University's (ANU) ePress initiative for electronic publishing advocates had created the environment which led to a bid under the A$250 million Systemic Infrastructure Initiative programme. This programme aims to: develop and document best practice; address strategic infrastructure issues; ensure solutions fit the Australian context; stimulate and share experiences. The purpose of the APSR project is to move repositories out of the development phase to become part of the research infrastructure.

The bid for the Australian Partnership for Sustainable repositories (APSR) was successful and in 2004 it was awarded a contract, to focus on an open standards based, long term sustainable, national programme to develop a range of repository-based services and to assist with this the project created a temporary repository of 5,000 papers. Within a year ANU had moved to evolve the development work into an operational and supported university service based on DSpace. In the larger community DSpace repositories now contain some 40,000 items and the development unit has 6 staff.[1] This success story does appear to rest on the ability of open access advocates, not simply to win the argument, but to align open access with larger funded agendas, where they are then seen as part of the solution to a wider agenda. But even with central funding, the issue

of advocacy remains very real. The 37 national repositories in Australia average just over 1000 articles each [12].

Netherlands

The Dutch experience is fully described in a recent article [10]. The Dutch research community, in this case championed by the IT community led by SURF, also developed a national strategy involving all thirteen universities and three major academic institutions, along with the national library. It has bid against and worked with the government's National Action Plan electronic highway. The focus has been on creating a consistent but not strait-jacketing infrastructure and aiming at coherence and interoperability, rather than completeness of deposit. The clear aim is to showcase research and the DARE Project appears to be very cleverly using academic vanity to encourage deposit, as well as having a large advocacy programme based on inclusiveness of stakeholders. The average number of articles in Dutch repositories is about 12,500 [12]. The project has been imaginatively extended by giving prominence to "more than 200 prominent scholars" who have been invited to showcase their publications on the website in the so-called Cream of Science [11].

Scotland

Following this analysis of national initiatives, OA advocates concluded that the problem of embedding change should be tested at national level. Although politically part of the United Kingdom, recent changes in devolved government have allowed Scotland to explore its traditional values and to gain much more control over its own future. The overarching government agenda is to make the country a hub in the global knowledge economy. National traits, political and social culture are then helpful in developing an Open Access strategy in Scotland and map neatly on to many of the arguments which support open access. [4]

There is a reverence for education, innovation and research. The country is small, with a population of five million people, which means that all interested parties can be brought together in a culture

G. Sica (ed.) *Open Access: Open Problems*

where working together is the norm. There is a tradition of social democracy (for further information see http://gdl.cdlr.strath.ac.uk/redclyde/) and a strong sense of community. There is also a clear recognition that as a small country, investment has to be made shrewdly and the results of that investment maximised. As in almost all small countries, pragmatism is valued at least as highly as principle, but at the same time there is a strong anti-establishment streak, making the Open Access agenda a natural issue for Scotland to support

The Scottish government agendas are also highly relevant. As stated above government is trying to position the country at the heart of the knowledge economy (Smart Successful Scotland) based both on inward investment, on research and on lifelong learning. Each of these demands access to up-to-date research and information for sustainable competitive advantage. With the worst health and dietary record in Western Europe there are major concerns over both social inclusion and health. Much of the research in these fields is commissioned by government, which wishes to see the research outcomes widely and freely available. Like many other west European countries, Scotland has a declining population resulting from a fall in the birth rate and a brain drain of the best and the brightest talents to other larger countries. Great importance is then attached to publicising and making public research which will show those outside the country the quality of research, thereby encouraging inward investment and to using repositories as a shop window for local researchers, encouraging them to stay in Scotland by demonstrating that major research opportunities exist at home. Finally, government is investing heavily in a programme called Digital Scotland, which is seen as providing the infrastructure which can underpin the issues above by delivering seamless access to a range of e-services

It is then a relatively straightforward process to map the open access agenda on to Scottish government agendas and demonstrate a range of potential benefits which coincide with the Open Access agenda. Thus IR advocates and government have a common ambition to demonstrate:

G. Sica (ed.) *Open Access: Open Problems*

- The distinctive nature of Scottish education and Scottish universities
- A desire to showcase an impressive research capacity – with 8% of the UK population Scotland wins 12% of the UK research awards
- Government awareness of the value of knowledge and access to it, with institutional repositories as the vehicle for marketing Scottish research
- The importance of a quality kite mark (peer review) and branding – research/knowledge products are branded as the output of the Scottish knowledge economy
- how to achieve "Best Value" – to modernise through e-government and broad use of e-service delivery
- the impact of Freedom Of Information legislation – moving towards a culture of access to information across a range of areas, especially in relation to public access to publicly funded research

In a small country politicians, government ministers and senior civil servants are accessible in a way that is not true of larger countries. It is hackneyed but true that everyone is related, or went to school or university together or supports the same football team. Promoting cultural change then becomes much more an outcome of personal persuasion than winning hearts and minds through logical argument.

Individual Scottish institutions had been involved in open access research and experimentation for some time. Various initiatives have established repositories across a number of Scottish institutions, providing the framework for a distributed, yet nationally co-ordinated approach working through a number of projects: HaIRST, Daedalus, Electronic Theses, Theses Alive, Oaisis.

But the collective journey towards open access in Scotland then began in October 2004 with the Scottish Open Access Declaration which was launched at an event at the Royal Society of Edinburgh attended by representatives of government, research funders, researchers, universities and librarians. The Declaration itself built on

G. Sica (ed.) *Open Access: Open Problems*

the growing number of declarations, particularly those of Budapest and Berlin and was quickly signed by all fifteen Higher Education institutions. Working with the government funded Scottish Library and Information Council, the research library community then focused around a project to develop a repository infrastructure. Beyond that project, advocacy continues at a personal level to have OA adopted as government policy.

3. IRIScotland: Institutional Repository Infrastructure for Scotland

The project has been set up with a view to addressing the issues of cultural change. It seeks to learn from experience in other small countries and to support the research agenda at both national and institutional level. The philosophy of research in Scotland is based on "pooled research". That is to say that in a small country which could sustain perhaps only one truly world class university, it is better to bring together the best researchers in a discipline, irrespective of their parent body with the aim of creating world class research in a discipline rather than a single institution.

The project then has three aims (JISC, 2005):

1. To explore ways of bringing about cultural and organisational change working with university senior managers and researchers to help in developing institutional research publication policies, procedures and mechanisms; to develop workflows to assist individual researchers which are conducive to the promotion of self-archiving in institutional repositories;
2. To develop a broad framework for a distributed institutional repository infrastructure for Scottish research and experiment with both a collective hosting repository, in particular for smaller institutions that may not wish to set up their own institutional repositories, and a cross-repository search facility capable of dealing with a wide range of research and research-related digital objects;
3. To identify what can be more effectively done centrally – and whether this should be done at a national Scottish level or a

> national UK-wide level – or locally at institutional level, taking account of relevant international developments to ensure that the Scottish infrastructure is globally interoperable.

In essence this will establish a consistent and standardised national network of repositories, which meet interoperable metadata standards, including a repository in the national library which will allow small research institutes to participate without setting up their own. This in turn will allow federated searching of all public sector research conducted in the country and provide a national shop window. At the same time work continues to lobby the Scottish Executive to mandate OA publishing of all publicly funded research and the vital work of advocacy in encouraging and handholding researchers through to deposit continues

In sum, having examined the issues which have delayed the universal acceptance of Open Access, its proponents in Scotland have felt that to address advocacy to the individual or to organisational structures does not address the problem of embedding cultural change. The proposed solution is then to attempt to map change on to national characteristics and government agendas. The nature of Scotland – its size and traditions – have been used to establish a national OA strategy, to be co-ordinated through the funded IRIS project. It also promises to further enhance the profile of Scottish research and thereby to deliver a number of crucial government agendas. Finally the use of international standards allows for interoperability and discovery

Conclusion

Small countries are the ones where information is on the agenda as a national priority. Where governments see or can be persuaded of the role of self-archiving through institutional repositories as a tool to leverage progress on other government agendas, notably skills retention and inward investment, there is a greater chance of piggy-backing a repository programme as an element of larger infrastructural programmes , rather than arguing for them as a good thing in their own right. Governments also tend to be major funding agencies – or at least important ones – and so are more

G. Sica (ed.) *Open Access: Open Problems*

amenable to recognising the logic and benefits of mandating deposit. Such countries also tend not to have large publishing industries (pace the Netherlands), so there is less incentive for the publishing industry to intervene and therefore a greater chance to promote sensible debate on the future of scholarly communication and less on the attack or defence of vested interest..

Bibliography

[1] Australian National University http://wiki.tertiary.govt.nz/~InstitutionalRepositories/Main/AustralianNationalUniversity.

[2] Directory of Open Access Journals http://www.doaj.org/.

[3] IRIS Project http://www.jisc.ac.uk/index.cfm?name=project_iriscotland.

[4] Law, Derek *et al* [2005] Developing a national information strategy in Scotland *Cadernos BAD* No 1. pp49-63.

[5] Lederberg, Joshua [May1996] Options for the Future *D-Lib Magazine*, http://www.dlib.org/dlib/may96/05lederberg.html.

[6] Moore, N. [2000] The international framework of information policies, in Elkin, J. & Law, D. (ed), *Managing Information,* Open University, Milton Keyne.

[7] New Zealand: http://www.digitalstrategy.govt.nz/templates/Page____11.aspx.

[8] Royal Society Response to Research Council UK's consultation on access to research outputs November, 2005 http://www.royalsoc.ac.uk/.

[9] Swan, Alma, Brown, Sheridan [2004] Authors and open access publishing. *Learned Publishing* 17(3), pp. 219-224.

[10] Van der Kuil, Annemiek, Feijen, Martin [2004] The Dawning of the Dutch Network of Digital Academic REpositories (DARE): A

Shared Experience *Ariadne* 41, http://www.ariadne.ac.uk/issue 41/vanderkuil/.

[11] Van der Kuil, Annemiek, Feijen, Martin [2005] A Recipe for the Cream of Science: Special Content Recruitment for Dutch Institutional Repositories *Ariadne* 45, http://www.ariadne.ac.uk/ issue45/vanderkuil/.

[12] Van Westrienen, Gerard, Lynch, Clifford A. [2005] Academic Institutional Repositories: Deployment Status in 13 Nations as of Mid 2005 *D-Lib Magazine* September.

Derek Law
Turnbull Building
University of Strathclyde
155 George Street
Glasgow G1 1 RD

Unbinding knowledge: a proposal for providing open access to past research articles, starting with the most important

Peter Suber

Abstract. If an authoritative scholar or organization assembled a bibliography of the most important previously published research articles on a subject of urgent public need, such as the treatment and prevention of HIV/AIDS, then the journals publishing those articles might be persuaded to provide open access to them retroactively. This article discusses the costs and benefits to journals participating in such a project, and calls on scientists, publishers, public-interest organizations, and foundations to experiment with such projects in order to accelerate research on topics where it is most needed.

1. Imagining a way to accelerate research

Most open access projects focus on new literature and leave open access to previously published literature an open problem for the future. This made sense in the early days of the open access movement, when helping hands, funding, and acceptance were all less common than they are today. Providing open access to new research articles is generally easier and less expensive than providing it retroactively to older articles. New articles are born digital; copyright holders are available for consent or persuasion; the benefits of increased audience and impact are most compelling at the time of publication; and a subsidy from the author's funding agency or employer is often available at that time and not later. There are still good reasons to make new literature a higher priority than older literature. But today, when open access has significant

G. Sica (ed.) *Open Access: Open Problems*

support and momentum, there are good reasons to include older literature in our strategic vision.

Some universities are asking their faculty members to deposit all their research articles, new and old, in the institutional repository. Some journals are digitizing their back runs and providing open access to them. PubMed Central will digitize and provide open access to the back runs of selected medical journals, and the Wellcome Trust spearheads a similar project.[1] These efforts are all critical and should continue. But many journals cannot afford to digitize their back issues; many are not eligible for PMC or Wellcome digitization (if only because they're not in biomedicine); and, most importantly, many are still *unwilling* to offer open access to their back issues. For literature in those journals, here's a step in the right direction.

The basic idea is for an authoritative scholar or organization to compile a bibliography of the 500 most important previously published research articles on a subject of urgent public interest, such as the treatment and prevention of HIV/AIDS. Then a hard-working soul asks the journals that published those articles to provide open access to them retroactively.

Let's say that the person compiling the list of articles is the *bibliographer*. It won't be hard to find a credible bibliographer. It could be a scholar with a track record in the field or a respected organization like the International AIDS Society.

Let's say that the person who contacts the journals is the *facilitator*. The facilitator might be a volunteer, but I expect that the work will be extensive and that foundations will be willing to fund it. The facilitator will have many tasks. First ascertain which articles from the bibliography are already open access and temporarily put them to one side. Then organize the remaining articles by journal. Contact each journal, identify the articles it published from the list, and ask it to provide open access to those articles. As needed, explain the request, answer questions, and negotiate details.

Some journals will accept these open access invitations and some will not. When the responses are in, the bibliographer and facilitator, or the organizations sponsoring them, will publicly thank the participating journals as well as the providers of the articles that were already openly accessible. The sponsors will produce a revised online

G. Sica (ed.) *Open Access: Open Problems*

version of the bibliography with links to open access editions of its articles.

Articles published in journals that decided not to cooperate will not appear on the list. The project purpose is to open the literature and applaud cooperation, not to publicize or shame non-cooperation. (More on this in Section 3, below.)

For lack of a better term, let's call such initiatives *unbinding projects*. I'm hoping that unbinding projects will be tried first in medicine, where the public good is easily recognized, the need is urgent, and open access is already familiar. However, the strategy is general and could be applied wherever there is clear social utility in accelerating the pace of research. Open access to the most important papers on artificial photosynthesis, fuel-efficient engines, or pollution-scrubbing smokestacks will advance research and development on these beneficial technologies.

2. Why would a journal agree?

Most journals receiving these requests will not be open-access journals, but most will have electronic editions. Providing open access to the articles identified by an unbinding project will not require peer review, editing, or manuscript preparation. It will only require moving a copy of the electronic file to an open-access web site.[2] At the same time, the journal might add some boilerplate text about the article's open-access status, and add an icon or annotation to the journal's table of contents to indicate that the article is now openly accessible.

Participating journals will help a good cause. They will help understand, treat, and prevent HIV/AIDS. To doubt this one must believe that good science is ineffective, that communicating scientific results does not make them more useful, or that the best scientific literature is already accessible to all those who can make use of it.

Participating journals will generate good will for themselves. They will become more visible as journals that published landmark articles. They will also become visible as journals willing to share knowledge and accelerate research on a matter of public importance. It shouldn't take an expensive marketing department to convert this

G. Sica (ed.) *Open Access: Open Problems*

kind of reputation into advantage in the competition for submissions, advertising, and subscriptions.

What about unbinding proposals on non-medical topics? Will journals feel the same moral force behind these requests? We won't know until we try the experiment, but it seems safe to assume that the journals publishing important articles on (say) wind power will be the same journals most receptive to the message that accelerating research on wind power is an important public good.

Peer review, editing, and manuscript preparation cost money, of course, but for almost every article included in an unbinding project, the journal will already have recovered these costs. Even if the journal sells access to its back run, this revenue is usually a very small part of its overall revenue stream. Moreover, participating in an unbinding project will only open up access to handful of its articles, not to the entire back run. The project does not ask a journal to accept or publish new articles, but merely to lift the protective barrier from previously published and previously amortized articles so that more people can make use of them.

The journal needn't release the articles into the public domain. If it is the copyright holder, then it can retain copyright.

Finally, the journal needn't worry that this kind of open access, even if extensive, would cause subscribers to cancel. That may be a worry with a general policy to provide open access to all new articles, or all past articles after an embargo period, but it does not arise with the selective, unpredictable, and retroactive opening of access to articles that turn out to be landmarks in their field.

It would be wonderful if all journals agreed to this offer, giving researchers, physicians, and the general public open access to 100% of the identified articles. But we can't expect that and we don't need that in order to justify the effort of trying.

Now imagine that similar bibliographies are compiled, and similar requests made, by the American Cancer Society, the Alzheimer's Association, the Royal Blind Society, the Epilepsy Foundation, the Leukemia and Lymphoma Society, the Malaria Foundation International....[3]

G. Sica (ed.) *Open Access: Open Problems*

3. Applauding, not shaming

Although the purpose is to open access to important science, and to applaud cooperation rather than shame non-cooperation, some important science journals may feel that a decision not to cooperate will be accompanied by a stigma. Working scientists will know that any list of important articles in their field will include some from that journal; hence, the journal's omission will be a sign of non-cooperation.

I see no way to avoid this perception. It will exist no matter how scrupulous the sponsors are to avoid any criticism of journals that do not cooperate. Different journals will give this perception different weights, but care and honesty in running the program will not eliminate it. Moreover, because this perception may be intrinsic to the decision not to cooperate, there may be no special reason to counter it.

But here are a few possibilities for countering it in case they are attractive to journals. All articles on the original list of 500 could be included in the published bibliography. This will give due recognition to the scientists who did the important work and the journals that published it. The drawback to this plan is that the bibliography could not link to open-access versions of articles from non-cooperating journals, and that could highlight their decision not to cooperate even more than omitting them. Another possibility is for the bibliography to link to priced or toll-access copies of these articles, at the journal web sites, in order to facilitate access for those readers who are willing to pay or who happen to have access privileges through their institutions. But these links would have to be labelled as toll access (or conversely, the rest would have to be labelled as open access) in order to help the sponsors identify the journals they wish to applaud for their cooperation. Since the balance of advantages and disadvantages for non-cooperating journals is a close call, the choice about these options might be left to the journals themselves.

A third possibility is to encourage the authors of the articles published in non-cooperating journals to provide open access to their articles themselves, through an institutional or disciplinary archive or their personal web site. Then the bibliography could link to that

G. Sica (ed.) *Open Access: Open Problems*

open-access copy. The drawback is that some journals still do not consent to postprint archiving. However, some journals that might not provide their own open-access copy of an article might consent to postprint archiving by the author.[4]

4. How much would this cost?

The sponsoring organizations would pay for identifying the most worthy 500 articles, discovering which are already available in open-access form, contacting the journals publishing the non-OA articles, making the request, explaining it, and conducting any follow-up negotiations. The sponsoring organizations would also bear the cost of preparing and hosting an online version of the resulting bibliography with active links.

All these costs could be paid by the same organization, or they could be split roughly as suggested between a disease advocacy organization (compiling the bibliography) and a research or humanitarian foundation (facilitating the rest). This is a natural opportunity for collaboration between scientific and philanthropic organizations and for cost-sharing between organizations with common interests.

The journals would pay the cost of copying a few files to a new location, adding boilerplate open-access licensing information to each one, and making minor revisions to a few tables of contents. When the electronic file is already prepared and the server costs are already part of the journal's overhead, then there is zero marginal cost in providing worldwide open access to the file, i.e. letting more people view the file that has already been created and is already online. As noted, the journals would already have paid the costs of soliciting, vetting, editing, preparing, and publishing electronic versions of the articles, and these costs were already recovered from the journal's subscription and licensing revenue.

Journals might be very proud that their articles were identified in the bibliography, and proud to enhance their usefulness by providing open access to them. If so, then they might take on the small additional cost of a press release, or even a press conference, to announce the unbinding.

G. Sica (ed.) *Open Access: Open Problems*

If the journal sells access to its back run, then it might lose something by providing open access to a very small but very important fraction of that back run. If it offers pay-per-view access to individual articles in the back run, rather than all-or-nothing access to the back run, then it may lose revenue from some of its best-selling articles. This is a real cost that should not be denied, although its amount will be difficult to ascertain and almost certainly small. A journal will naturally take this cost into account when deciding whether to agree to the unbinding request.

Because journals need not worry that this form of selective, unpredictable, retroactive open access will cause subscribers to cancel, their decision on the unbinding proposal will turn almost entirely on lost revenue from future sales of access to the identified articles and whether this is outweighed by the resulting benefits for public health and the journal itself. Since typically the lost revenue will be vanishingly small, the public good large, and the good will generated for the journal considerable, we can predict that many journals will agree to participate.

The costs to the sponsoring organizations will certainly be greater than the costs to any single journal (from unbinding), and probably greater than the unbinding costs of all the participating journals combined.

5. Moral suasion v. money

Because a journal's costs in complying with the unbinding request are so low, it is reasonable to ask journals to make this gesture for the sake of assisting an important cause. If moral suasion does not work and the sponsoring organizations have the means, then journals could be paid to release the selected articles into open access.

Moral suasion is preferable to money for two reasons. First, participating in an unbinding project costs a journal very little. Second, the sponsoring organizations would not have to pay an extra cost to bring about this public good, freeing them to pursue similar projects on other fronts and do more for the public good with its limited resources.

The publishers participating in the HINARI program, for example, are providing free online access to entire nations in the

developing world without asking philanthropists to pay their costs. They are producing the journals in electronic form anyway. Giving access to new readers who couldn't buy subscriptions costs the journals nothing, generates good will, and enfranchises huge segments of humanity. It's definitely win-win.

Someone might object that the HINARI analogy isn't entirely apt because some of the researchers who would read the articles on the HIV/AIDS bibliography would be in a position to pay for access. That is true. But extending the same line of thought shows how to answer the objection. Many of the researchers, perhaps most of the researchers, who would want to read and use the articles on the bibliography will be affiliated with institutions that already subscribe to the relevant journals. They're not just in a position to pay; they're already paying. So instead of losing more revenue, the journal gets the best of both worlds: good will for providing open access and paid access by most researchers at the same time.

Another reason why moral suasion might work here is that the journal approached to unbind a given article is the only journal that can make this particular contribution to the public good. That makes the unbinding request very different from the request for a monetary donation, no matter how good the cause. Money is fungible and anyone else could give it. This well-known fact dilutes the moral force of the request and leads many good people to hope that their neighbors will make up for them. By contrast, scholarly journals are notoriously non-fungible. This fact is normally an obstacle to open access: if journals were fungible, then journals in the same research niche would compete directly with one another, and (assuming comparable quality and prestige) affordable journals would kill off the expensive ones, and free journals would kill off everything else. But an unbinding project turns the non-fungibility of journals to advantage. People asked for help are most likely to give it when they are in a unique position to do so. When a journal is asked to provide open access to one of its articles, then it can't assume that it was targeted in bad faith (as if the article were really unimportant) or that anyone else could provide this public service.

There's no reason to rule out the possibility that some requests will be accompanied by money and some will rely on moral suasion alone. The two can coexist. If the only way to free up the most

G. Sica (ed.) *Open Access: Open Problems*

important articles on superconducting ceramics, Dutch elm disease, or human-powered flight is to pay for them, then possibly someone could be found who would pay for them.

6. Variations on the theme

- Free up the 500 most important papers on HIV/AIDS in 2006. Two years later liberate the most important 100 published during those two years. Repeat every two years or until all new research is open access.
- To save time and money, the group funding the bibliographer could start with existing bibliographies in the field.
- To increase the acceptance and authority of its bibliography by researchers in the field, at the expense of some time and money, the group funding the bibliographer could poll experts in the field, or even assemble an eminent editorial board just for this purpose. It could also post the draft bibliography online for a comment period.
- Journals that appreciate the logic of this project might spontaneously offer open access to articles in their back run that have somehow been identified as important. For example, when a scientist receives an award for a research breakthrough, a journal publishing some of his or her past articles could instantly provide open access to them. At the same time, it could issue a press release announcing that these important articles are now free to the research community and general public.[5]
- Likewise, individual scientists who appreciate the logic of this project might ask the journals that published their articles on matters of public importance to provide open access to them retroactively, or even immediately. Journals might be more responsive to organized projects than individual initiatives, but a groundswell of individual requests might be as effective as an organized project. Neither scientists nor journals need wait for brokers or philanthropists to mediate the unbinding requests.

- Participating in an unbinding project, or simply witnessing other journals participate, will demonstrate how open access can benefit journals and publishers, and not just their authors and readers. By making a journal's brand more visible, and cementing its reputation as a journal publishing important articles and serving important public needs, participation can help a journal compete for the next generation of important articles, not to mention advertising and subscriptions. As journals learn more about these benefits, some will move other articles from their toll-access back run to their open-access back run. Some will realize that they gain more from increasing the visibility and accessibility of their articles, at least after a certain time, than from selling access to them.
- There is clearly nothing magical about the number 500. It would be better for HIV/AIDS patients if all articles useful for treatment and prevention were openly accessible than just 500. And on some new or narrow topics, there may be far fewer than 500 useful articles already published. The bibliographer can set this number in accordance with the topic and size of the body of published literature. When scientists bypass the bibliographer and facilitator, and take their requests directly to journals, they will clearly have no particular number in mind, and journals shouldn't feel constrained by a number.
- The sponsors could create a wide-ranging web site devoted to their unbinding project. In addition to the bibliography with active links, and perhaps the archive of papers, the site could contain the sponsors' press release of public thanks and some background on the project's purpose and method. If the project's beneficiaries (for example, article authors, other researchers in the field, and AIDS patients, their families, friends, and physicians) want to offer public thanks and testimonials, this web site would be the natural place to do so. The site could also provide information for donors willing to help offset the project's costs.

G. Sica (ed.) *Open Access: Open Problems*

- When one or two of these projects have been brought to a successful conclusion, advocacy organizations will be eager to propose new projects and scholars will be eager to nominate articles for them. If the early projects are difficult, we can expect some of the difficulties to disappear as the idea becomes more familiar and we gain experience in implementing it.

7. An objection from the side of open access

It would be ironic if subscription-based journals liked this idea more than open-access proponents did. Open-access proponents might object that this is not true open access, and that by satisfying journals, authors, and readers, it might delay progress toward true open access.

Here's my quick reply to the objection. As the Bethesda Statement on Open Access Publishing makes clear,[6] open access is a property that belongs primarily to individual articles, and only derivatively or secondarily to journals. An unbinding project will create true open access to the articles from participating publishers. What an unbinding project will not do, on its own, is produce open-access journals that provide open access to all their articles past and future. That is true. But while there are many, highly diverse strategies for persuading conventional journals to offer open access to more of their articles, unbinding projects would not interfere with any of them. On the contrary, the most promising strategy I see for getting subscription-based journals to consider open access seriously is to get them to experiment with it. Unbinding projects are among the easiest and least risky kinds of open-access experiment. Moreover, they do not hinder the launch of new open-access journals or the spread of open-access archiving, and even boost them by directly acquainting more authors, readers, journals, and publishers with the benefits of open access. Finally, while other open access strategies are at work, having their own effects at their own speeds, this one could be enlarging the body of open-access literature, focusing on previously published articles that would otherwise remain behind access barriers, and starting with those that could be most helpful to urgent scientific and social problems.

G. Sica (ed.) *Open Access: Open Problems*

Conclusion

Opening access to important research articles will accelerate research and all the benefits of research, from new medicines and therapies to improved clinical practice. It will also benefit the journals that published the articles by generating good will, increasing their visibility, and enhancing their reputation for scientific excellence and humanitarian assistance, all of which can translate in to bottom-line advantages in the competition for submissions, advertising, and subscriptions. This method of opening access can harness the interests, energy, and resources of groups that normally have little involvement in scholarly publication. It costs journals very little. Many journals will find that it brings them a net gain and many will consent even in the face of a small loss because of the gain it brings to others. Unlike other open-access initiatives, which focus on future literature, this one opens up past articles, starting with the most useful and important. It introduces journals to the methods and economics of open-access publishing, which could lead to more experimentation with open-access publication. It doesn't ask journals to convert to open access, but limits the request in scope and risk, making it easier for journals to assess and accept. It is a frank business proposition, with true benefits for the journal to weigh against the costs. It invites deliberation, not confrontation, and moves the open-access question from sometimes obstreperous conferences to the quiet of the journal's business office. Finally, it is likely that many journals will see it as a win-win proposition, agree to it whole-heartedly, and thereby enlarge the body of open-access research literature, make their own important articles more useful, and accelerate research on a matter of vital public need.

.....

I thank Barbara Cohen, Helen Doyle, and Debra Lappin for comments on an earlier draft of this article. I give special thanks to Darius Cuplinskas for the stimulating conversation in which this idea emerged almost fully formed. He is certainly its coauthor.

.....

Notes

[1] For some of the best university policies encouraging or requiring faculty to deposit their research output in the institutional repository, see the Institutional Self-Archiving Policy Registry.
http://www.eprints.org/signup/fulllist.php
(accessed November 16, 2005)

Unfortunately there is no definitive list of non-OA journals with OA back runs. But a good collection of such journals is available at Highwire Press.
http://highwire.stanford.edu/
(accessed November 16, 2005)

PMC Back Issue Digitization Project.
http://www.pubmedcentral.gov/about/scanning.html
(accessed November 16, 2005)

Medical journals backfiles digitisation project, sponsored by the Wellcome Trust, the Joint Information Systems Committee (JISC), and the National Library of Medicine (NLM).
http://library.wellcome.ac.uk/node280.html
(accessed November 16, 2005)

[2] There are several ways to do this. The journal could move a copy of the electronic file to a newly-created open-access directory on its own web site. An open-access directory is simply one that is not password protected. Or the journal could deposit a copy of the file in PubMed Central or another open-access repository for that discipline. Since both acts are trivial, and take only minutes of someone's time, a journal could do both.

The leading open-access publishers today do both. BioMed Central and the Public Library of Science host their own open-access copies of their articles, and deposit copies in PubMed Central.

BioMed Central
http://www.biomedcentral.com/
(accessed November 16, 2005)

Public Library of Science
http://www.plos.org/
(accessed November 16, 2005)

PubMed Central
http://www.pubmedcentral.nih.gov/
(accessed November 16, 2005)

[3] For a partial list of disease advocacy organizations, see the membership roster of the U.S.-based National Health Council, http://www.nationalhealthcouncil.org/aboutus/membership_index.htm

[4] "Preprint archiving" is the deposit of an unrefereed preprint in an open-access archive. "Postprint archiving" is the deposit a refereed version, or a version approved by a journal's peer-review process.

For a searchable database of publisher policies on preprint and postprint archiving, see Project SHERPA, http://www.sherpa.ac.uk/romeo.php

[5] When a scientist wins the Nobel Prize, clearly the journals that published his or her work would benefit science as well as their own standing if they provided open access to the breakthrough articles. But the stimulus might come from a turn of events, not just the rising star of an author. For example, after the September 11 attacks, McGraw-Hill offered open access to a full-text book, Glenn R. Schiraldi's *The Post-Traumatic Stress Disorder Sourcebook: A Guide to Healing, Recovery, and Growth*. The fact that McGraw-Hill was helping victims of post-traumatic stress disorder and advertising itself at the same time did not undercut the usefulness of its action. On the contrary, it is precisely by mixing self-interest with public service that public service becomes more likely and more sustainable.

[6] See the Bethesda Statement on Open Access Publishing, June 20, 2003, http://www.earlham.edu/~peters/fos/bethesda.htm
(accessed November 16, 2005)

Peter Suber
Open Access Project Director, Public Knowledge
Research Professor of Philosophy, Earlham College
Senior Researcher, SPARC

G. Sica (ed.) *Open Access: Open Problems*

The emerging role of open repositories for scientific literature as a fundamental component of the public research infrastructure

Paul F. Uhlir[1]

Abstract. It is axiomatic that the rate of change in technological systems initially outpaces human capacity to adapt to the technological advances. This is particularly true for transformational technologies that displace their antecedents, including the associated organizational paradigms. Such a transformation is currently taking place as a result of the technological revolution brought about by the combination of digital information technologies and global communication networks. Among the unresolved tensions resulting from this paradigm shift in the information sector is between the publishers who have controlled the means of production and distribution of scientific information in the print medium and those seeking to reorganize the institutional frameworks and business models to take advantage of the unprecedented, transformational attributes of digital networks.

This article argues that the legacy business model for research journals that restricts access to and use of publicly funded scientific information is not taking advantage of the potential benefits of global open availability online and is therefore slowing the progress of

[1] The views expressed in this article are those of the author and not necessarily those of the National Academies. The author gratefully acknowledges the support of the National Library of Medicine at the National Institutes of Health under contract number 467 MZ 501 180.

G. Sica (ed.) *Open Access: Open Problems*

science and socioeconomic development. The question, therefore, is not whether open availability to such publications is better than access provided on terms that are economically, legally, and technologically restricted, but how open availability can be most effectively institutionalized and how quickly. Nevertheless, it is not desirable to destroy the legacy model without operationally sustainable alternatives that can provide open availability.

There are, in fact, multiple paths that are being taken to make this transition from conditionally restricted to open dissemination. Some of these, such as open access publishing, are beginning to replace the legacy model, whereas others, such as open repositories for scientific literature, have been shown to co-exist with subscription-based publishing, albeit with substantial unease. The evolutionary human systems are thus beginning to respond to the opportunities made possible by the revolutionary technologies embodied in global digital networks. A gradual, but highly significant and far-reaching restructuring of scientific communication in public research is taking place. This process needs to be better understood and managed.

The article describes the reasons why the publications produced through publicly funded research must be made openly available online, reviews some of the emerging models for providing open availability to them, and discusses the role of open institutional repositories for such publications as a fundamental component of the public research infrastructure. The article concludes with some thoughts about possible actions that might be taken by the various stakeholders working in this arena to accelerate this transition.

1. Introduction

It is axiomatic that the rate of change in technological systems initially outpaces human capacity to adapt to the technological advances, much less to exploit those advances for maximum social and economic benefit. This is particularly true for transformational technologies that displace their antecedent technologies and the associated organizational paradigms. In such cases, not only do new management approaches need to be adopted in response to technological progress, but they must overcome the substantial

G. Sica (ed.) *Open Access: Open Problems*

resistance to change by the entrenched interests whose business model is based on the superseded technology.

Such a transformation is currently taking place as a result of the technological revolution brought about by the combination of digital information technologies and global communication networks. Each of these technologies represented a significant paradigm shift in its own right. Communication satellites in the 1960s accelerated the global integration process initiated by the development of telephone networks decades earlier, presaging the advent of the "Global Village". The relatively rapid introduction of digital technologies and their subsequent connection through communication networks, most notably the internet, has resulted in profound social and economic transformations in many countries and sectors. Few information products or aspects of human activity have remained untouched by this technological revolution. These transformations, which are now commonly referred to as the "Information Age", the "Information Society", or the "Knowledge Economy" are only now becoming more broadly understood.

This article looks only at one small slice of this transformational process—the effect of digital networks on the distribution and use of the peer-reviewed literature and related information, specifically in publicly-subsidized research journal[2] publishing. Box 1 presents a comparison of the characteristics of publishing under the print paradigm with those of global digital networks[3].

The magnitude of the changes that are possible from the shift from print to digital technologies and networks cannot be overstated, either quantitatively or qualitatively. The explosion in the production

[2] The term "research journals" in this article refers to scientific, technical, and medical (STM) journals.

[3] Based on a summary of a presentation by Uhlir, Paul F., "Re-intermediation in the Republic of Science: Moving from intellectual property to intellectual commons," in *Open Access to Scientific and Technical Information: State of the Art and Future Trends*, Herbert Gruttemeier and Barry Mahon, eds., IOS Press, International Council for Scientific and Technical Information, Paris, 2003, pp. 63-66.

G. Sica (ed.) *Open Access: Open Problems*

of digital bits is now well known as a function of Moore's law[4]. Among the other well-known quantitative advantages that digital networks have over the previous print paradigm are in time, geographical extent, and cost; that is, digital networks can provide instantaneous, concurrent, and global availability at near-zero marginal cost of access by each additional user. These quantitative improvements make possible, if not yet realized, the universal availability of information.

Box 1: Comparison of some key characteristics of the print and digitally networked paradigms

PRINT	GLOBAL DIGITAL NETWORKS
■ (pre) Industrial Age	post-industrial Information Age
■ fixed, static	dynamic, interactive
■ rigid	flexible, extensible
■ physical	"virtual"
■ geographically limited	global
■ linear	non-linear, asynchronous, with time/space collapsed
■ limited content and types	unlimited contents and multimedia
■ distribution difficult, slow	easy and immediate dissemination
■ copying cumbersome, not perfect	copying quick, simple, identical
■ significant marginal distribution cost	near-zero marginal distribution cost
■ single user (or small group)	multiple, concurrent users/producers
■ centralized production	distributed production
■ knowledge diffusion slow	accelerated diffusion, network effects

Just as important, however, are the qualitative advantages of digital technologies and networks in accelerating the dissemination of information and the diffusion of knowledge. Because networks provide the opportunity for non-linear, interactive, and asynchronous communication with multimedia capabilities, the potential to improve the dissemination and diffusion processes also has been greatly magnified. The digital nature of the information imbues it with flexible transformative properties, making it subject to easy manipulation and integration with other types of information to create new knowledge that was either not possible or

[4] Moore's law posits that the computing power of microprocessors doubles every 18 months. See, e.g., http://en.wikipedia.org/wiki/Moore's_law.

much more difficult in the print context. Moreover, the network enables entirely new forms of collaborative knowledge production on a broadly distributed, interactive, and even anonymous basis, changing or dis-intermediating the hierarchical and centralized organizational models through which information was produced and knowledge diffused in previous eras[5]. And, perhaps most important, digital networks make possible entirely automated approaches to information extraction, processing, integration, and organization of vast amounts of information, which can be transformed into unlimited new discoveries and products, eclipsing the capabilities of human information production, dissemination, and use.

As both the principal inventors and among the most pervasive users of the internet, scientists have a great deal at stake in fully exploiting the potential of this new medium for accelerating scientific progress and its benefits to society. These quantitative and qualitative features of the emerging digital networks are realized most completely, however, when the information is made freely and openly available online. Among the unresolved tensions in this context is between the publishers who have controlled the means of production and distribution of scientific information in the print medium (and now online) and those seeking to reorganize the institutional frameworks and business models to take advantage of these unprecedented, transformational attributes of global digital networks. Although the legacy organizations have begun to adapt to and adopt the new technological framework, they have not been leading the changes that would maximally exploit the potential of digital networks for the publicly funded research enterprise[6]. The

[5] For a seminal article on the institutional, economic, and legal aspects of the evolving volunteer, distributed, peer-production models online, see Benkler, Yochai (2002), "Coase's Penguin, or, Linux and the Nature of the Firm", 112 *Yale Law Journal* 369.

[6] For an excellent historical perspective, see Guédon, Jean-Claude (2001), "In Oldenburg's Long Shadow: Librarians, Research Scientists, Publishers, and the Control of Scientific Publishing", in ARL Proceedings 138, Association of Research Libraries Publications Distribution Center, 70 p., available at: http://www.arl.org/arl/proceedings/138/guedon.html. For a comprehensive, continually updated bibliography of information

initial *agents provocateurs* and innovators of such changes therefore have arisen almost exclusively from the research community itself—from the universities, libraries, and individual researchers—with growing support from public and private funding organizations.

This article describes the reasons why the publications[7] produced through publicly funded[8] research must be made openly available online, reviews some of the emerging mechanisms for providing open availability to them, and discusses the role of open institutional repositories for such publications as a fundamental component of the public research infrastructure. The article concludes with some thoughts about possible actions that might be taken by the various stakeholders working in this arena to accelerate this transition.

2. The Case for Open Availability of Publications from Public Research

The scientific process relies on open communication to make the results of prior work available for verification, analysis, and incorporation into new results. Over the centuries this process of

resources about open access publishing, see Bailey, Charles W., "The Open Access Bibliography", available at: http://www.escholarlypub.com/oab/oab.htm/.

[7] This article focuses exclusively on open repositories for scientific literature, primarily from peer-reviewed journals. It does not address open repositories for data, which are discussed in a separate, though related, article: Uhlir, Paul F., and Peter Schröder (forthcoming), "Promoting Access Publicly-Funded Scientific Data for Global Science", *Data Science Journal*, CODATA.

[8] The arguments presented here are most compelling in the context of scientific literature produced predominantly from public funding as a global public good. As discussed in Part II, however, the motivations and interests of authors who are funded by private-sector sources who publish their scientific results in scientific journals make the arguments presented here equally compelling to them and most likely to their employers or funding sources. Because the relationship between the authors who are funded predominantly by private sources and the journals in which they publish is essentially a private matter, rather than one of broad public concern, they are generally excluded from express consideration in this article.

G. Sica (ed.) *Open Access: Open Problems*

communication evolved through a system of journals and books published by university presses and learned societies, and supported by conferences and other types of meetings focused on the exchange of information and the diffusion of knowledge.

Following World War II, there was a great increase in both public and private funding of research and development in the industrialized countries, which spawned the creation of thousands of new outlets for communication, both peer-reviewed literature and more informal exchanges. The universities and learned societies were at the forefront of this proliferation of publications. Under the print paradigm, the only way to distribute these results was through the sale of books and through subscriptions to serial publications.

In more recent decades, an increasing percentage of these publications has been produced and disseminated by commercial publishers, who have viewed this system as a low-cost, high-profit business opportunity. In some cases, the learned societies have licensed their publishing operations to the commercial publishers, while in other cases the societies have maintained their publishing ventures in-house. Either way, the societies have used the publishing proceeds to support their journal operations and ancillary activities as not-for-profit entities.

It is becoming quite obvious, however, that the print subscription journal model for the communication of results from public research has become inadequate in the emerging era of global digital networks. The availability of information from publicly funded research on an open basis[9] provides substantially greater benefits on digital networks than a subscription model based on a restrictive access basis. Even more important, the increase in benefits is not linear, because of network effects. This is true from a scientific, economic, social, and ethical perspective, as discussed below.

[9] There are several different models for making information openly available, which are reviewed in subsequent sections. In this section, the discussion focuses solely on open versus restricted modes of dissemination, without factoring in the relative costs and benefits of different open access models.

G. Sica (ed.) *Open Access: Open Problems*

2.1 The scientific imperative of openness

The motivations of government, academic, and other not-for-profit researchers generally are not market-driven, but tend to be based more on non-economic, public interest factors. They are rooted predominantly in intellectual curiosity, the desire to create new knowledge, peer recognition, career advancement, and public service. They place a premium on having their ideas widely known and used. And, as authors, they are typically willing to pay to publish; certainly scholarly journal publishers never pay them for their contributions[10].

Research funding organizations and managers of the public research enterprise have a corresponding duty to the research community and to the broader public to promote a productive and sustainable research base, and to maximize the realization of the value of the public investment in that research. As the Research Councils of the United Kingdom (RCUK) recently noted, research funders and managers should seek to ensure that:

> Ideas and knowledge derived from publicly-funded research are made available and accessible for public use, interrogation, and scrutiny, as widely, rapidly and effectively as practicable. New internet-based models for the publication of research outputs and also of the underpinning data are likely to play an increasingly useful role in the widening and speeding of access, which in turn supports the Research Councils' strategies for ensuring that the results of research are exploited more effectively for the benefit of the UK's society and economy[11].

In light of the motivations and values of most researchers in the public (and frequently even in the private) sector, as well as of the managers and funders of public research, perhaps the greatest failings of the restricted dissemination model is the suboptimal access to and utilization of the research results. It is certainly well known that much of the published literature distributed under the

[10] See, generally, Berry, Steven R. (2001), Is Electronic Publishing Being Used in the Best Interests of Science? The Scientists' View, 2001 *International Journal of Molecular Sciences* 133.

[11] RCUK Position Statement on Access to Research Outputs (28 June 2005), p. 1, available at: http://www.rcuk.ac.uk/access/statement.pdf/.

G. Sica (ed.) *Open Access: Open Problems*

restricted subscription model has gone unused, unread, and un-cited in subsequent research. This unfortunate situation was understandable and tolerable when the only means of dissemination was in the print format, where only a few journals were broadly distributed and read and most were – and could only be – made available in very limited quantities. Moreover, the proliferation of titles made the competition for readership intense and made it unlikely or difficult for scholars to obtain and use all the available information relevant to a particular research problem.

Digital networks have the potential to change this underutilization dramatically, but only if access to the information is unimpeded by artificially imposed economic, legal, and technological barriers[12]. Open availability in the online environment creates an unobstructed, instantaneous, global, and potentially comprehensive database on the desktop of each researcher. It greatly enhances and facilitates: scientific inquiry, diversity of analysis and opinion, new types of research, and methods of analysis[13], and prevents duplicative work and research inefficiencies generally[14]. It eliminates the cost barriers on the transfer of information to researchers and students who are unable to afford high subscription prices, or the costs and delays of interlibrary loans. Overall, the different forms of open access online promote interdisciplinary and international research, particularly in

[12] Recent studies have shown an increase of 2 to 5 times in citation impact and broad visibility and use of material made openly available versus through the restricted subscription approach. See, e.g., Harnad, *et al*. (2006) "The green and the gold roads to Open Access", *Nature Web Focus* 18 February 2006, citing also several other studies, available at: http://www.nature.com/nature/focus/accessdebate/21.html; see also, Hajjem, Chawki, *et al*. (2005) "Ten-Year Cross-Disciplinary Comparison of the Growth of Open Access and How it Increases Research Citation Impact", *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering* 28(4), p. 39-46..

[13] Arzberger, Peter, *et al.*, "An International Framework to Promote Access to Data," *Science* 303(1777), 19 March 2004, discussing the benefits of open access to data from publicly funded research.

[14] Committee on Information, Computer and Communications Policy (2005), *Digital Broadband Content: Scientific Publishing*, Organisation for Economic Co-operation and Development, Paris, p. 69.

integrating scientists in the developing world more into the global research system, and accelerate scientific progress and innovation by making diverse information resources much more easily available. In this context, it should be noted as well that scientific breakthroughs do not necessarily come from the linear application of scientific knowledge, but result instead from leaps of knowledge discovery, frequently aided by serendipitous considerations triggered by a perusal of the literature.

A variety of technological advances for automated literature-based discovery[15] also continue to improve the potential accessibility and automated utilization of information online, but the full benefits of such advances in information retrieval and manipulation can only be realized if the material is openly and fully searchable. Automated software applications for a broad range of processing and analysis are constantly improving. If the full text of the papers could be indexed by computer programs one could navigate through them automatically in support of a range of information extraction and re-aggregation functions, as long as appropriate references maintain reputational benefits. One early example of this is CiteSeer, which has over 700,000 papers with full text in the computer sciences and has automatically created a very useful digital library and research resource. The purpose of CiteSeer is "to improve the dissemination and feedback of the scientific literature and to provide improvements in functionality, usability, availability, cost, comprehensiveness, efficiency, and timeliness on the access of scientific and scholarly knowledge[16]". Instead of creating another digital library, the project provides algorithms, metadata, software, and other services that can be used in other digital libraries freely and openly. Nevertheless, programs such as CiteSeer are limited in their potential utility

[15] See, e.g., Kostoff, Ronald N. (2005), "Systematic Acceleration of Radical Discovery and Innovation in Science and Technology", *Storming Media*, available for a fee at: http://www.stormingmedia.us/02/0270/A027034.html/; and Berkman, Paul A., *et al*. (publication pending), "Automated Granularity to Integrate Digital Records: The 'Antarctic Treaty Searchable Database' Case Study", *Data Science Journal*, CODATA.

[16] See http://citeseer.ist.psu.edu/.

G. Sica (ed.) *Open Access: Open Problems*

because of restricted journal access and the threat of intellectual property liabilities.

Perhaps most obviously, there is less well informed research possible if the publicly-funded information is not made freely and openly available. High access costs result in significant lost opportunity costs that are certain to occur, but are difficult to measure. If information is too costly or difficult to obtain, the potential user – whether a researcher or some other individual – will opt to use a different source or will work without this information, limiting the breadth of analysis and diversity of opportunities that might otherwise be realized. Overall, the overprotection and high costs of access to the public research results taxes the innovation system in each country and slows scientific progress[17]. As the recent RCUK Position Statement noted, "there is widespread concern…that the full potential of the internet for enabling members of the research community as a whole to gain access to journal articles on their desktop is not being exploited; and that researchers in many institutions cannot gain access to the full range – and the full runs – of the journals they need. This is clearly a concern to the Research Councils…."[18]

2.2 The economic benefits of openness

The economic benefits of open availability of publicly funded scientific information may be assessed at several levels – in terms of the relative economic efficiency of the dissemination function, in better capturing the value of the information, and in promoting the network effects and related positive externalities associated with its open availability on the internet.

It is important at the outset to understand the public good characteristics of publicly funded research and the information

[17] For a discussion of these issues in the context of publicly-funded research data see: Reichman, JH, and Paul F. Uhlir (Winter/Spring 2003), "A Contractually Reconstructed Research Commons for Scientific Data in a Highly Protectionist Intellectual Property Environment", in *Law and Contemporary Problems*, Vol. 66, Duke University School of Law, at p. 410-416.

[18] RCUK, *op. cit.*, note 11.

produced through it. Unlike private goods, public goods have non-excludable and non-rival characteristics.[19] The former means that once a good has been produced, the producer cannot exclude others from benefiting from it. The latter attribute is that it costs nothing to provide the good to any additional person, which means that the marginal cost is zero. A lighthouse, despite its high initial production cost, is typically cited as a pure public good. Once operational, no ship within its visible range can be excluded from its benefit and the benefit provided by the lighthouse costs exactly the same (although its value can vary substantially) whether only one ship or a thousand sail within its ambit.

Although there are few such pure public goods, there are many "quasi"-public goods that substantially possess both non-excludable and non-rival characteristics. The national defense is another activity that is frequently described as a public good. Once constituted, it is difficult (though not impossible) to exclude some citizens and not others from the protection, assuming that it is truly national in scope. Similarly, a comprehensive national defense also is non-rival, although one may argue that increasing defense funding will in fact help assure more complete coverage for the entire population.

Both basic scientific research (without any near-term prospects for exploitation in commerce) and the information derived from it also may be considered quasi-public goods.[20] Ideas, facts, information, and knowledge, once divulged, may be freely communicated among individuals and are difficult to keep from others. Both are non-rival, although they are, to a certain degree, appropriable and excludable. Their excludability, however, must be artificially created and enforced[21]. The exclusive restrictions result in

[19] See, e.g., Inge Kaul *et al*. (1999), "Defining Global Public Goods", in *Global Public Goods: International Cooperation in the 21st Century*, Kaul *et al*., eds.

[20] Callon, Michael (1994), "Is Science a Public Good?", *Science, Technology, and Human Values*, Vol. 19, p. 400.

[21] In the case of journal publishing, such artificial exclusion is created through the assignment of exclusive copyright held by the publicly funded authors to the journal publishers, and subsequently enforced by the publishers against the public research community through legal and, in digital form, technological means.

G. Sica (ed.) *Open Access: Open Problems*

what economists refer to as a "deadweight loss" from sub-optimal utilization[22].

Most types of public research – for example, on tropical diseases, astronomy, and climate change, or the large facilities or spacecraft used for their study – fall within this category. Their high costs and poor commercial prospects make public investment the only means by which to support them. Yet their overall contribution to the "knowledge infrastructure[23]" and spin-off applications make them socially beneficial activities, sufficiently attractive to garner public support[24].

Similarly, the data and information produced through public research form part of this knowledge infrastructure. Because science builds on science, open communication of the results and the maximum diffusion and dissemination of that information is the guiding principle and principal interest of the public research system. The value of this information lies in its use, not in the capture of maximum rents by monopolies in an artificially created market. To the extent that the information is fixed in tangible form, however, it is potentially excludable, even if exclusion does not promote the overall objectives and values of the public research system.

Under the print paradigm, each user (or group of users via a library or institutional copy) of a journal represented an additional cost of dissemination to the publisher and therefore the copy did not have non-rival characteristics, as well as being potentially excludable. Because of the substantial costs involved in producing and distributing the print copies, a subscription-based system was deemed to be a reasonable self-supporting way to disseminate the information in a sustainable way.

[22] See Scotchmer, Susan (2003), "Intellectual Property: When Is It the Best Incentive Mechanism for S&T Data and Information?", in *The Role of Scientific and Technical Data and Information in the Public Domain*, Julie M. Esanu and Paul F. Uhlir, eds. National Academies Press, pp. 15-18.

[23] David, Paul (2001), "The Political Economy of Public Science", in *The Regulation of Science and Technology*, Helen Lawton Smith, ed.

[24] See generally, Reichman & Uhlir, *op. cit.*, note 17.

The journal system under the print paradigm essentially had no alternative but to treat the information output as a quasi-private good, even if the "market" in which it operated was almost entirely an artificial one created and subsidized by public funds. The information published in the journals was submitted freely to the publishers (or even with page charges, to further defray publishing costs) because its production was fully subsidized by the authors' public research grant or employing institution. Much of the labor used to maintain quality control – the peer review or editorial service – was also provided freely or on a low-cost basis. And the subsequent acquisitions of the literature by libraries and individual researchers were fully or partially subsidized through public research and education funds. An artificial monopoly was granted to the publishers by the authors' assignment of copyright, even though the information was created through a public good activity and the authors received no compensation for the copyright transfer[25].

The publishers, for their part, added value (and still continue to do so) to this otherwise heavily subsidized enterprise in many ways. They vetted the submitted manuscripts and prepared the ones that

[25] As analysts at Credit Suisse First Boston succinctly noted:

> [W]e would expect governments (and taxpayers) to examine the fact that they are essentially funding the same purchase three times: governments and taxpayers fund most academic research, pay the salaries of the academics who undertake the peer review process and fund the libraries that buy the output, without receiving a penny in exchange from the publishers for producing and reviewing the content…..
>
> We do not see this as sustainable in the long term, given pressure on university and government budgets.

In *Sector Review: Scientific, Technical and Medical Publishing*, April 6, 2004 (quoted from a presentation by Heather Joseph, "Open Access Initiatives: Public Access Policies in the United States", International Seminar on Open access, 9th World Congress on Health Information and Libraries, Salvador, Brazil, September 21, 2005). Indeed, many government policy makers and research funders (as well as researchers and librarians) have begun to question the sustainability and logic of the legacy restricted access publishing model, as discussed further below.

G. Sica (ed.) *Open Access: Open Problems*

were reviewed and accepted for publication by providing editing, fact checking, layout, and other production functions associated with quality control. They also disseminated the copies of the works and promoted their availability through marketing and targeted advertising.

Initially, the system was administered almost exclusively by the learned societies on a not-for-profit basis, so that the deadweight social costs of this artificial monopoly were minimized. The publishing activity served both the scientific community and the public interest. In recent decades, however, and still before the advent of pervasive digital networks, the benign economics began to be exploited much more vigorously by commercial publishers, who saw a low-risk, high-value profit center in this artificial market. At the same time, many of the learned societies have either transferred their publications for a fee to the commercial publishers or used their subsidized monopolies to generate excess revenues to support their other activities.

In short, it should not be surprising that information protected by a monopoly status and maintained on a closed proprietary basis, particularly in a market dominated by a few large companies, will tend to result in higher, anti-competitive pricing. The commercialization of public research publishing has led to what are by now well-known, massive increases in journal serial subscription rates, which have far outstripped both the consumer price index over the past two decades[26]. Moreover, managing publicly funded information on a restrictive, proprietary basis has added substantial administrative complexities on both ends of each transaction, including subsequent enforcement of the proprietary restrictions, further taxing the public research system with avoidable overhead costs.

Digital networks have changed the economic dynamics of this increasingly dysfunctional arrangement. The publishers must still

[26] See Association of Research Libraries (2003), "Monograph And Serial Costs in ARL Libraries, 1986-2002", in *ARL Statistics 2001-02*, Washington, DC., showing the subscription costs paid by libraries to publishers exceeding the growth of the consumer price index by approximately 150 percent.

recover their production costs, of course. However, the social costs associated with restricted distribution models when the marginal cost of dissemination online is practically zero make it essential to adopt new or complementary models based on the open availability of this public good. Because of the pervasive access to the internet by researchers in all but the poorest countries and institutions (which eventually will become part of the network as well), the maximum economic value from public or publicly funded scientific information can be derived from its free and open availability (the marginal cost of dissemination online)[27], and encouragement of its re-use on a permissive licensing basis that seeks only to protect the reputational benefits of the authors in their publications.

Most important in the digital network context, however, are what economists call "positive externalities" and "network effects". An externality is the action of one party affecting the well-being of another, without adequate compensation[28]. An externality may be positive, where such an action confers an uncompensated benefit, or negative, where it imposes additional costs on the other party(s). Basic research and its publicly funded information outputs may have no direct economic applications, but frequently are credited with leading to serendipitous advances, innovation, and commerce. Such benefits are clearly much more likely to occur in an open information dissemination system, where access and use are maximized, rather than in a restricted access system[29].

[27] Stiglitz, Joseph A., *et al*. (2000), "The Role of Government in the Digital Age", Computer & Communications Industrial Association.

[28] Ibid.

[29] For some (imperfectly) analogous studies of the relative economic and social benefits of open versus closed information regimes in the public sector, see European Union (1999), "The European Commission's Green Paper, Public Sector Information: A key resource for Europe", annex 1 at 20-25 COM 585, available at: http://europa.eu.int/ISPO/docs/policy/docs/COM(98)585/; and PIRA International (2000), "Commercial Exploitation of Europe's Public Sector Information, Final Report for the European Commission", Directorate General for the Information Society, available at: http://europa.eu.int/information_society/policy/psi/docs/pdfs/commercial_exploitation/commercial_final_report.pdf#search='pira%20international%20commercial%20exploitation'/.

Network effects multiply the potential benefits from open availability and the related positive externalities exponentially. Communication networks, including the internet, derive exponential value from the addition of individual users to the network. And, as discussed in the preceding sub-section, the value of vast quantities of information online that can be extracted most efficiently automatically with new software tools that dwarf even the greatly increased potential of individuals using the network. This greatly under-exploited opportunity to improve the extraction of value from online databases and text is perhaps the main reason over the longer term to make as much of the publicly funded scientific information as openly available as possible.

2.3 Social and ethical values of openness

A final strong argument in favor of open availability of scientific information produced through public funding revolves around the social benefits and related ethical dimensions. Because both the research and the information produced as a result of it are publicly funded, public-interest, public goods, the public maintains an interest in seeing that the benefits are maximized and the costs minimized. The scientific and economic benefits of openness discussed above are certainly socially beneficial.

The social and ethical aspects have both general considerations and those that are more specific to certain kinds of information. As a general matter, the public has already paid for the production of the information, but the burden of fees for access falls disproportionately on the poorest and most disadvantaged institutions and individuals. The continually rising and frequently exorbitant subscription fees for scholarly journals increasingly force cancellations at many university libraries, diminishing the investment value of the information and effectively placing those institutions in a second-rate status (or ensuring that they stay there). As one university professor in the United States put it recently:

> I teach at one of the universities that has been marginalized by this process. In this new information age, at my university, our ability to access the scholarly literature actually is decreasing each year as our library is forced to subscribe to fewer and fewer electronic

> journals. In addition, unlike paper journals where we still had a copy on the shelf, what we had access to last year is now gone…. Should we simply abandon scholarly work and research in our poorer states? That is, should we allow the marketplace to determine which universities will have access to the core research literature and which will not?[30]

But if major universities in the United States are cancelling significant numbers of serial subscriptions, the situation can be truly dire in less economically developed countries, where the subscriptions are almost entirely unaffordable[31]. Unlike users in the best financed institutions in developed countries, such structural inequalities make it impossible for researchers, educators, and students (as well as practically anyone else) to access and use the world's knowledge base to help improve their R&D and educational capacities. Moreover, just as high-priced and artificially restricted information from public research widens the knowledge gap between the rich and the poor, restricted access publishing models in developing countries similarly tend to render much of that STM literature invisible in the South and greatly limit the potential readers worldwide. This makes it difficult for researchers in the developing world to actively integrate into the global research system and for research institutions in the South to utilize or even track the research output in their own countries[32].

[30] Onsrud, Harlan (2003), "Emerging Models for Maintaining Scientific Data in the Public Domain", in *The Role of Scientific and Technical Data and Information in the Public Domain*, Julie M. Esanu and Paul F. Uhlir, eds. National Academies Press, p. 180.

[31] It should be noted that free or heavily subsidized access to research journals has been made available to many developing countries through several major programs, including the International Network for the Availability of Scientific Information (PERI), the United Nations World Health Organization (HINARI), and the United Nations Food and Agriculture Organization (AGORA). Many individual publishers and research institutions also now provide such preferential access in developing countries to their otherwise fee-based, restricted materials.

[32] See Chan, Leslie, Barbara Kirsop, and Subbiah Arunachalam (2005), "Open Access Archiving: the fast track to building research capacity in developing countries", *SciDevNet*, available at: http://www.scidev.net/ms/openaccess/.

Access fees, of course, can pose a significant if not insurmountable barrier to obtaining the scientific literature by individuals outside the research system as well. As noted above, open availability of the scientific literature maximizes use, but not just within the research community. Both the citation rates and the readership base of many open access journals has been shown to be greatly expanded, and this has been shown to be the case in the context of some developing country journals[33]. This is particularly important in subject areas of great public interest, such as biomedical advances, socioeconomic studies, environmental research, and space exploration, to note but a few. The constraints on access from high subscription fees in publicly funded research journals, which otherwise constitute a (global) public good, thus raise important ethical and social considerations that have begun to elicit the attention of policy makers and legislators in some countries, as discussed below, and of other stakeholders in the research community worldwide.

* * *

The artificially imposed excludability in information produced from publicly funded research (and in other public information goods) through the use of high access costs, maximum intellectual property restrictions, and electronic fences subverts all these values and potential scientific and socioeconomic benefits of global open availability online. The question, therefore, is not whether open availability to such publications is better than access provided on economically, legally, and technologically restricted terms, but how

[33] For one such focused case study, see Greene, Lewis Joel (2005), "Effect of Open Access on a Scientific Journal Published in a Developing Country" (presentation given at the International Seminar on Open Access, 9th World Congress on Health Information and Libraries, Salvador, Brazil, 20-23 September 2005), available at: http://www.icml9.org/meetings/openaccess/public/documents/7%20ap%20-%20Lewis%20Greene-185422.pdf/.

open availability can be most effectively institutionalized and how quickly[34].

Nevertheless, although the business model under the print paradigm was based almost universally on a subscription fee model, it is not desirable to destroy that legacy approach without an operationally sustainable alternative that can provide open availability[35]. There are, in fact, multiple paths that are being taken to make this transition from restricted to open dissemination, not just for publicly-funded[36] scientific literature, but for other types of information as well. Some of these, such as open access publishing, are beginning to replace the legacy model, whereas others, such as open repositories, have been shown to co-exist with subscription-based publishing, albeit with substantial unease.

The evolutionary human systems are thus beginning to respond to the opportunities made possible by the revolutionary technologies embodied in global digital networks. A gradual, but highly significant and far-reaching restructuring of scientific communication in public research is now taking place. This process needs to be better understood and managed. Indeed, as the RCUK has pointed out: "It is likely that for the foreseeable future, many different models and

[34] The same question may be asked about access to data from publicly funded research. See, Uhlir and Schröder, *op. cit.*, note 7.

[35] This point, understandably, is made most forcefully by the legacy publishers who rely on the subscription-based business model. See, e.g., The Royal Society of the United Kingdom (2005), "The Royal Society position statement on 'open access'", available at http://www.royalsoc.ac.uk/page.asp?id=3882.

[36] This article focuses on the rationale and models for providing open availability to the peer-reviewed journal literature derived from publicly funded research. The focus is on the publication of publicly funded research results because of the inherent public-interest aspects that are subject to public policy considerations and control by public research institutions. As noted in Part II, scientists from private-sector institutions who publish in science journals have the same or similar interests to see their ideas and writings made as openly available as possible. However, the institutional context is different and not subject to the same rationale and solutions as in the public sector, although a similar strategy could be devised for the private sector.

G. Sica (ed.) *Open Access: Open Problems*

mechanisms will co-exist, and…[i]t may even be that increasing development of the internet will render the traditional "publication" of journal articles less relevant or prevalent as a means of communicating research results[37]." The next section provides an overview of the new approaches to producing and disseminating public research results that have emerged over the past decade or so as part of this research restructuring process.

3. Emerging Models for Open Availability of Scientific Information

In light of the clear benefits to the research enterprise and to society from open availability of publicly-funded scientific information in the digitally networked environment, it is not surprising that a variety of new models already have been developed within the research community. Table 1 provides a summary of the different types of open access approaches for diverse scientific information resources.

Table 1: Summary of open access models for different types of scientific information

There are many new kinds of distributed, open collaborative research and information production and dissemination on digital networks.
Examples of open data and information production activities include: ➢ Open-source software (e.g., Linux and thousands of other programs worldwide, many of which originated in academia[38]); ➢ Distributed Grid computing (e.g., SETI@Home[39]);

[37] RCUK, *op. cit.*, note 11.

[38] The Free Software Foundation directory lists over 4300 free software packages at : http://directory.fsf.org/.

[39] The Search for Extraterrestrial Intelligence project was the first to use the parallel processing power of many thousands or even millions of distributed, individual computers on a volunteer basis, linked through the internet, to help solve complex computing-intensive problems. See: http://setiathome.berkeley.edu/. For a comprehensive list of dozens of such projects in science and other fields of activity, see:

- Community-based open peer review (e.g., Journal of Atmospheric Chemistry and Physics[40]); and
- Collaborative research Web sites and portals (e.g., NASA Clickworkers[41], Wikipedia[42], Project Gutenberg[43]).

The following are examples of open data and information dissemination:

- Open data centers and archives (e.g., the U.S. National Library of Medicine's National Center for Biotechnology Information[44], the Protein Data Bank[45], SNP Consortium[46], Digitized Sky Survey[47]);
- Federated open data networks (e.g., World Data Centers[48], Global Biodiversity Information Facility[49]; NASA Distributed Active Archive Centers[50]);
- Virtual observatories (e.g., the International Virtual Observatory for Astronomy[51], Digital Earth[52]);
- Open access journals (e.g., *BioMed Central*[53], *Public Library of Science*[54], + > 2100 scholarly journals[55]);

http://distributedcomputing.info/. A useful overview of the distributed computing activities may be found in a special issue on Distributed Computing in *Science* 308(5723), 6 May 2005, p. 809-824.

[40] See: http://www.copernicus.org/EGU/acp/.

[41] See: http://clickworkers.arc.nasa.gov/top/.

[42] See: http://en.wikipedia.org/wiki/Main_Page/.

[43] See: http://www.gutenberg.org/info/volunteer/.

[44] See: http://www.ncbi.nlm.nih.gov/.

[45] See: http://www.rcsb.org/pdb/.

[46] See: http://snp.cshl.org/.

[47] See: http://archive.stsci.edu/dss/.

[48] See: http://www.ngdc.noaa.gov/wdc/guide/wdcguide.html/.

[49] See: http://www.gbif.net/.

[50] See: http://nasadaacs.eos.nasa.gov/.

[51] At this point there are only national or regional astronomy portals that have not yet been integrated. For an overview of the planned International Virtual Observatory Alliance, see: http://www.ivoa.net/pub/info/.

[52] See: http://www.digitalearth.gov/.

[53] See: http://www.biomedcentral.org/.

[54] See: http://www.plos.org/.

[55] A registry of scholarly open access journals is maintained by the Lund University Libraries in Sweden. See: http://www.doaj.org/.

- Open access hybrids, based primarily on a subscription model, but including a mix of restricted subscription-based and open author-pays articles (e.g., *PNAS*[56], *Springer Open Choice*[57])
- Delayed open availability, following some restrictive period that varies from one or more months to one or more years (e.g., *PNAS* or *Science*[58], which make their subscription based articles available after six months and one year, respectively);
- Open institutional repositories for an institution's scholarly works (e.g., the Indian Institute for Science,[59] + > 400 globally[60]);
- Open institutional repositories for publications in a specific subject area (e.g., PubMedCentral[61] for biomedical literature, or the Cornell arXiv[62] for *e*-prints in physics, mathematics, nonlinear sciences, computer science, and quantitative biology);
- Free university curricula online (e.g., MIT's OpenCourseWare[63]); and
- Emerging discipline-based commons (e.g., the Conservation Commons[64]).

The common elements of all these different types of initiatives are that the information is made openly and freely available digitally and online. In many cases, the material is either made available under reduced proprietary terms and conditions through permissive licenses[65] (e.g., the GNU license for open source software[66], or

[56] See *Proceedings of the National Academy of Sciences*: http://www.pnas.org/.

[57] See: http://www.springer.com/sgw/cda/frontpage/0,11855,1-40359-12-115382-0,00.html.

[58] See: http://sciencemag.org/.

[59] See: http://eprints.iisc.ernet.in/.

[60] A registry of open access institutional repositories is maintained by Southampton University in the UK. See: http://archives.eprints.org/eprints.php?action=home/.

[61] See: http://www.pubmedcentral.nih.gov/.

[62] See: http://arxiv.org/.

[63] See: http://ocw.mit.edu/OcwWeb/.

[64] See: http://conservationcommons.org/.

[65] For an overview of such permissive licensing approaches spanning all information types, see Liang, Lawrence (2005), "A Guide to Open Content Licenses", Piet Zwart Institute, Rotterdam, available at: http://pzwart.wdka.hro.nl/mdr/research/lliang/open_content_guide/.

Creative Commons licenses[67] for open access journals or for some works in open repositories), or it is in the public domain[68] . In other cases, such as delayed open availability by STM publishers to their journal articles, the works remain protected under full copyright, but are freely and openly accessible.

Just as the desirability of providing open availability to publicly funded scientific information online was substantiated in Part II above, the many different models that have already been established attest to the feasibility of doing so. The various examples now provide valid proofs of concept for all information types, in most disciplines, in many countries, and across all institutional categories – whether government agencies, universities, not-for-profit organizations, or even for-profit firms.

Taken together, these activities are part of the emerging broader movement in support of both formal and informal peer production and dissemination of publicly funded scientific (and other) information in a globally distributed, volunteer, and open networked environment[69]. They are based on principles that reflect the cooperative ethos that traditionally has imbued much of academic and government research agencies. Their norms and governance mechanisms may be characterized as those of "public scientific information commons," rather than of a market system based upon

[66] See http://www.gnu.org/.

[67] The Creative Commons organization develops standard “permissive use” licenses with “some rights reserved”. Creative Commons recently launched the Science Commons, a project focused in part on licenses that promote greater access to and use of scientific data and information. See: http://science.creativecommons.org/.

[68] Information is in the public domain when it meets the following conditions: (1) the information is not copyrightable, such as factual compilations or data sets that lack creativity and originality in their selection and arrangement; (2) the information is produced by a government that does not apply copyright to its own works (e.g., the United States federal government); or (3) the statutory period of intellectual property protection has expired, which in many jurisdictions now is the life of the author plus 70 years.

[69] Benkler, *op. cit.*, note 5.

G. Sica (ed.) *Open Access: Open Problems*

proprietary data and information[70]. Such information commons activities respond – either explicitly or tacitly – to the scientific, economic, and socio-ethical factors discussed in Part II.

The journal literature forms a large and important subset of this emerging global information commons for public science. Policy principles and declarations in support of open access to the publicly funded journal literature have been issued in several countries since 2002 and have involved many institutions[71]. Legislation in support of more open availability of the literature has been introduced in the United States[72] the United Kingdom[73], and, most recently, Germany[74].

[70] David, Paul A., and Paul F. Uhlir (2005), "Creating the Information Commons for *e*-Science: Toward Institutional Policies and Guidelines for Action", CODATA Newsletter 91.

[71] See, e.g.: the 2002 Budapest Open Access Initiative, available at: http://www.soros.org/openaccess/read.shtml/; the 2003 Bethesda Statement on Open Access Publishing, available at: http://www.earlham.edu/~peters/fos/bethesda.htm/; the 2003 Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, available at: http://www.zim.mpg.de/openaccess-berlin/berlindeclaration.html/; the 2004 Messina Declaration, issued by 32 rectors of Italian Universities, confirming their adherence to the Berlin Declaration, available at: http://www.aepic.it/conf/viewappendix.php?id=49&ap=1&cf=1; the 2004 Finnish national commitment to open access publishing, available at: http://www.news-medical.net/?id=1411; the 2004 Scotland declaration in support of open access, available at: http://scurl.ac.uk/WG/OATS/declaration.htm/; the 2005 United Kingdom Research Councils' position statement and recommendations, available at: http://www.rcuk.ac.uk/access/index.asp/; the 2005 U.S. National Institutes of Health statements on its open access policy, available at: http://publicaccess.nih.gov/; and the 2005 Salvador [Brazil] Declaration on Open Access: the developing world perspective, available at: http://www.icml9.org/channel.php?lang=en&channel=91&content=439/. For a comprehensive listing that is continually updated, see Suber, Peter, "Timeline of the Open Access Movement", available at: http://www.earlham.edu/~peters/fos/timeline.htm.

[72] The first such legislative initiative in the United States, the "Public Access to Science Act", was introduced in June 2003 in the U.S. House of Representatives. Although it contained some problematic provisions and was not acted upon, it established an initial legislative marker and had important symbolic value. Among the bill's provisions were a declaration

Other countries are now actively reviewing their national policies in this area[75]. These public policy initiatives are indicative of a growing consensus within the international scientific community that open models of access to public research results, rather than economically,

that "U.S. government-funded research belongs to, and should be freely available to, every person in the U.S.", and a requirement that "…any Department or Agency should make every effort to develop and support mechanisms for making the published results of research…freely and easily available to the scientific community, the private sector, physicians and the public."

The following year, the U.S. House of Representatives Committee on Appropriations in its Fiscal Year 2005 Report Language stated that: "The Committee is concerned that there is insufficient public access to reports and data resulting from NIH-funded research", and that this situation was "….contrary to the best interests of the U.S. taxpayers who paid for this research."

The CURES Act was introduced in the U.S. Senate on 7 December 2005, which would, among other provisions, require open access to federally funded medical research publications after a 6-month embargo period. See: http://lieberman.senate.gov/documents/bills/051207curessectionbysection.pdf/.

Most recently, the Federal Research Public Access Act of 2006 was introduced in the U.S. Senate, extending the requirement of open access to publications from federally-funded research, not just medical research. The Act would apply to all agencies that fund extramural grants totaling more than $100 million per year. See: http://cornyn.senate.gov/doc_archive/05-02-2006_COE06461_xml.pdf /.

[73] In 2004, the U.K. House of Commons Science and Technology Committee recommended that all U.K. higher education institutions establish institutional repositories on which their published output can be stored and read online, free of charge.

[74] The German Bundesrat (the Upper Chamber of the Parliament) introduced legislation in May 2006 that would permit self archiving of journal articles by authors six months after publication, regardless of the terms they may have signed in any copyright transfer agreement. See, Suber, Peter, "Top stories from May 2006", in the SPARC Open Access Newsletter, No. 98, June 2, 2006, available at: http://www.earlham.edu/~peters/fos/newsletter/06-02-06.htm#topstories/.

[75] Countries that are now actively considering the adoption of open access policies at the national level include Australia, Finland, South Africa, and Sweden. *Ibid.*

G. Sica (ed.) *Open Access: Open Problems*

legally, and technologically restricted ones, are essential. There are two main roads that lead to the broad open availability of this body of knowledge – one taken by publishers and one by the authors[76]--as outlined below.

3.1 Open access publishing and hybrid models

Publishers have adopted several approaches to provide open availability to the publicly funded scientific literature. The new model that enables the most comprehensive and immediate access at reduced levels of proprietary restrictions is, of course, the "open access" journal. According to the Bethesda Principles, an open-access journal publication is one that meets two conditions:

1. The author(s) and copyright holder(s) grant(s) to all users a free, irrevocable, worldwide, perpetual right of access to, and a license to copy, use, distribute, transmit and display the work publicly and to make and distribute derivative works, in any digital medium for any responsible purpose, subject to proper attribution of authorship[2], as well as the right to make small numbers of printed copies for their personal use.
2. A complete version of the work and all supplemental materials, including a copy of the permission as stated above, in a suitable standard electronic format is deposited immediately upon initial publication in at least one online repository that is supported by an academic institution, scholarly society, government agency, or other well-established organization that seeks to enable open access, unrestricted distribution, interoperability, and long-term archiving (for the biomedical sciences, PubMed Central is such a repository)[77].

[76] These are now commonly referred to as the "gold" and "green" roads to open access, as coined by Harnad, S., *et al.* (2004), "The Access/ Impact Problem and the Green and Gold Roads to Open Access", Serials review 30, available at: http://eprints.ecs.soton.ac.uk/9939/

[77] The Bethesda Statement, *op. cit.*, note 71.

Open access journals are funded by a mix of sources that may include external institutional grants, internal institutional support, author submission fees or page charges, advertising revenues, and various donations of facilities, equipment, and labor. As noted in Table 1, there are now over 2100 scholarly journals that have self-registered as operating under open access conditions on the internet. New organizations have been formed to provide open access publishing services for client journals, either on a remunerated or subsidized basis. Some of the larger, more notable, examples include BioMed Central in the United Kingdom, and the SciELO[78] network and Bioline International[79] in developing countries.

In response to the surge in new open access journals, many legacy subscription publishers, especially the not-for-profit professional societies and even some commercial publishers, have begun to adopt new strategies or experimental policies that provide some form of open availability to their otherwise restricted, proprietary material. In 2004, nearly 70 professional society journal publishers jointly produced the "DC Principles," which also recognized the imperative of broad access to the scholarly literature produced from publicly funded research[80]. This statement, however, stopped well short of endorsing immediate open access models, promoting instead a voluntary delayed open access approach.

Under the delayed open availability scheme, publishers make their journal contents freely available online following some initial period of restricted access, typically six months or one year after initial publication. The publishers generally retain full intellectual property rights in the material and only allow access to the final published version through their official Web sites. This approach is an extension of price discrimination and product differentiation already used by the legacy publishers for the pricing of access to different customer groups, in this case dropping the cost of access to the marginal cost online following a time period in which most of the

[78] See: http://www.scielo.br/.
[79] See: http://www.bioline.org.br/.
[80] See: http://www.dcprinciples.org/.

G. Sica (ed.) *Open Access: Open Problems*

potential revenues have already been captured[81]. Nevertheless, there is a significant social cost associated with each delay in open availability that also must be considered.

Another interesting approach taken by a few legacy publishers is the hybrid access model. This allows an author to choose whether to pay a fee to obtain immediate open access to the author's article or to make it available under the standard subscription basis. Unlike the "pure" open access journal approach, however, the publishers using the hybrid model retain copyright and do not release the material under a Creative Commons license.[82] Moreover, only a few publishers have adopted this approach, although they have included both not-for-profit[83] and commercial[84] entities. In most cases, these experiments have been too recent to draw any firm conclusions about them. One such preliminary result reported by *PNAS* publisher, Ken Fulton, has shown that the "PNAS Open Access articles receive 50% more full-text accesses and PDF downloads than subscription-access articles in the first month after publication and maintain higher usage in subsequent months"[85].

3.2 Open access deposits by authors

Whereas open access publishing and variations thereof have been adopted by some publishers as part of their business model, self-

[81] This is based on what economists call "Ramsey pricing", which was developed for regulated monopolies. For a discussion of this economic principle and pricing strategies for the scientific data and information market, see National Research Council (1997), *Bits of Power: Issues in Global Access to Scientific Data*, National Academy Press, pp. 124-126.

[82] Creative Commons, *op. cit.*, note 67.

[83] The Proceedings of the National Academy of Sciences adopted the hybrid access model on a trial basis from January 2004 to January 2006. *Op cit*, note 56.

[84] Springer was the first commercial publisher to experiment with the hybrid access approach through its "Open Choice" program, which it initiated in August 2005. *Op cit*, note 57.

[85] Rowlands, Ian, and Dave Nicholas (2005), "New Journal Publishing Models: An International Survey of Senior Researchers", CIBER, available at: http://ucl.ac.uk/ciber/ciber_2005_survey_final.pdf/.

archiving online allows an individual author to provide open availability of a particular work. Self-archiving is an access mechanism rather than a publishing mechanism. There are essentially two types of outlets for authors to provide open availability for their formally refereed, peer-reviewed journal articles as well as their un-reviewed publications or "gray" literature: (1) open institutional repositories[86], and (2) open thematic repositories.

Open institutional repositories – whether based at a government agency, university, not-for-profit organization, or even a private-sector institution – can serve as a free digital collection of the works of an institution's employees and affiliated professionals. The main purpose of such repositories is to highlight and broadly disseminate the intellectual output produced under the auspices of their *organization*. There are now many thousands of these repositories in different institutions throughout the world, as more and more organizations realize the value of placing their output openly online. Southampton University in the UK maintains a registry of such repositories, which now numbers over 400, mostly at universities[87].

Open thematic repositories provide free e-prints of journal articles and other literature based on the *subject matter or discipline* of the information, rather than on the organizational affiliation. These materials may be limited to the final pre-publication versions of articles that have been reviewed and accepted for publication, or the publisher's version itself. They may also include pre-prints posted for

[86] Personal Web sites, particularly those affiliated with the author's institution of employment, may be considered as subsets of the institutional repositories, although there clearly can be some important differences in the way they are managed and especially archived. Such a discussion, however, is beyond the scope of this article.

[87] See: http://archives.eprints.org/eprints.php. An open repository must self-register to be included on this site, which is why there are only 400+ listed there, whereas the estimate of the actual number of such repositories, either in existence or under development, numbers in the thousands. For example, all Dutch universities and approximately 85 percent of German universities have set up institutional repositories, and it is expected that all United Kingdom universities will have institutional repositories established within the next few years. RCUK *op. cit.*, note 11.

public commentary prior to the article's formal publication. As in the case of the institutional repositories, thematic repositories are likely to contain other types of information together with the journal articles.

Understandably, there are considerably fewer thematic repositories than institutional repositories because there are not nearly as many disciplines or even sub-disciplines as there are institutions. One of the best known open thematic repositories is the Cornell arXiv[88] (formerly at Los Alamos National Laboratory), which was originally established in 1991 for pre-prints and post-prints in high-energy physics and now has been expanded to include several other fields of research. Another is PubMedCentral[89], which was created in 2000 for post-prints of the biomedical and life sciences literature. There are several other such open thematic repositories, including some more in the biomedical and life sciences, cognitive sciences, computing research, mathematics, chemistry, social sciences, and library and information sciences[90].

Although institutional and thematic repositories are not yet ubiquitous, they have established a legitimate proof-of-concept approach that merits a lot more attention. Moreover, most of these repositories have been established from the bottom up; that is, at the individual institutional level (including the thematic ones), rather than as top-down policy at the national level. Nevertheless, as noted above, national legislatures, science policy makers, and major research funders are now beginning to take a closer look at the functioning of the journal system, and at the complementary roles that open repositories can play in maximizing the value of the public investment in research. Given the foregoing, the time is now ripe for the public research community at all levels to take a much closer look at the relative merits of open repositories in the digital network context.

[88] Available at: http://arxiv.org.

[89] Available at: http://www.pubmedcentral.nih.gov/.

[90] The Association of Learned and Professional Societies Web site lists these subject-matter archives, available at: http://www/alpsp.org/htp _openarc.htm/.

4. Open Repositories as a Fundamental Component of the Public Research Infrastructure

Open institutional and thematic repositories already have been shown to serve a critical function in the digitally networked research enterprise, whether or not subscription-based publishing eventually is superseded by open access publishing. There are several reasons for this. First, placing the papers in an open repository is clearly in the author's interest. As noted in Part II, authors who are publishing the results of their research in the peer-reviewed journal literature are certainly not in this game for the profits they derive directly from their publications. The open availability of their research writings only enhances their professional interests and reputations, whether they are in public or private sector research.

Second, a large percentage of subscription-based publishers, both not-for-profit and commercial, already allow some form of immediate, post-publication self-archiving by their authors[91]. This means that the open availability of these works could be realized with only minimal delay if authors were to self-archive their works and they all had a suitable place in which to do so.

Third, and related to the preceding point, neither the open institutional nor the thematic repositories have been shown to undermine the revenue stream of the legacy publishers. This is true even in those cases, such as the physics arXiv, where the repository has been operating for 14 years in a relatively small and specific subject area. Unlike the open access publishing model, which requires a publisher to make wholesale conversion from one business model to another, with perhaps uncertain results, open institutional and thematic repositories thus far have co-existed with subscription journals without negative economic effects on the publishers (but with positive effects for science and society, as discussed in Part II).

[91] See current statistics at: http://romeo.eprints.org/stats.php/. According to this source, as of 5 June 2006, 69 percent of registered journals (6466) allowed post-print deposits by authors and an additional 24 percent (2248) allowed pre-print deposits.

G. Sica (ed.) *Open Access: Open Problems*

Finally, the open availability of the journal literature is not only in the authors' interest, but promotes the progress of science and is broadly in the public interest. Thus, even if open repositories were in some instances to reduce the subscription revenues of journal publishers (as long as such reductions did not destroy their financial viability), for all the reasons stated above, such repositories would still be broadly in the interest of science, society, and the economy. This point is clearly the most controversial one, at least from the perspective of the legacy publishers, but it flows logically from the arguments presented here.

In addition to these broad benefits of open repositories for digital peer-reviewed journal articles (and other information products) produced through publicly-funded research, there are more specific benefits associated with each of the two major open repository models.

4.1 Specific benefits of open repositories

Although both types of repositories clearly advance the interests of authors, institutional repositories tend to provide benefits to the organizations that operate them, whereas thematic repositories tend to be more useful to the discipline they serve and to the funders of the research.

For example, open institutional repositories at universities can provide a permanent and globally visible record of their achievements in research and teaching, allowing people both inside and outside the institution to discover the work undertaken there. This function can be enhanced by linking among the different types of information across the institution. Faculty can access and reuse these materials in teaching and future research. The open availability serves as an advertisement of the institution's capabilities and accomplishments, helping to attract new funding sources, new complementary faculty and research talent, and more focused student applicants. Higher citation rates and greater readership of the institution's scholarly output enhance the prestige of the employees and the organization. Finally, certain academic administrative tasks

can be facilitated, such as compiling lists of journal articles written by the staff and sorting them according to different criteria[92].

Open thematic repositories specifically enable improved access to discipline-related research, which the funding agencies find more useful to their interests. By aggregating articles from research that they have funded under a certain programmatic or discipline area, such an interoperable, thematic collection adds value to the individual works, with the value of the whole greatly exceeding the sum of its parts. It demonstrates accountability for the research funding that was made by the funding agencies and the taxpayers, and can help attract additional research proposals. As in the case of institutional repositories, the aggregation of these (thematically and programmatically) linked papers can assist the funding agencies (and especially researchers themselves!) in determining whether certain research has already been undertaken prior to funding similar new proposals[93]. Similarly, the peer review process can be facilitated if all the referenced papers are readily accessible through searchable archives[94]. Plagiarism and fraudulent results also can become much more easily detected and therefore less likely to enter or remain in the literature[95].

By integrating the journal articles with other thematically linked research output such as the underlying data sets, the gray literature, and multimedia products, and providing various software for information manipulation and interactive services that enhance the

[92] Friend, Fredrick J. (2005), "Why European universities and funding agencies are committing to open access", presentation given at the International Conference on Strategies and Policies for Open Access to Scientific Information, Beijing, China. See also OECD, *op. cit.*, note 14, pp. 67-69.

[93] *Ibid.*

[94] Chan, *et al.*, *op. cit.*, note 32.

[95] See Hilf, Eberhard R. (2003), Report on the IUPAP Workshop on Scientific Misconduct and the Role of Physics Journals in its Investigation and Prevention. A specific example of how this can work in practice is the CiteSeer project, which can show the similarity of all openly available papers in the computer sciences at the sentence level, as well as co-citation similarity and active bibliography. Personal communication from Lee Giles, 21 December 2005.

value of the repository to users, the repository can become vastly more useful for research, education, and other applications than a mere collection of articles in a single journal series[96]. The creation of such "knowledge environments" becomes greatly facilitated and achieves maximum network effects and beneficial externalities if all the thematically linked materials are openly available. One can thus envision a future in which there are not only multiple open thematic repositories covering many major areas of human knowledge, but also the possibility of individuals or small groups using new digital tools to create customized knowledge environments that can automatically extract and integrate vast amounts of disaggregated, but openly available and usable information resources to advance knowledge in ways that are now only dimly perceptible.

5. Clearing the Roadblocks for the Open Road

Up to this point, most of the discussion has focused on answering the question of why the publicly-funded journal literature should be made openly available on digital networks, and then why open repositories are becoming an essential component of the research infrastructure. We now turn to some of the other questions and issues about how this transition might be most effectively implemented.

Of course, scenarios of intellectual nirvanas in an open access universe, while serving as worthy longer-term goals, require a host of sometime difficult, incremental steps to be taken. The digitally networked revolution, with all its great promise and potential, will always be subject to evolutionary human policies and practices. Certainly, the transformational nature of the technology makes the human adaptation more complex. The possibilities for disintermediation of some of the legacy institutions that are unable or unwilling to adapt their business models in the greater service of public science generates additional resistance.

[96] See, e.g., Chan, L. (2004), "Supporting and Enhancing Scholarship in the Digital Age: The Role of Open Access Repositories", *Canadian Journal of Communications*, 29 (3-4), 277-300.

The community of practitioners and researchers who have been analyzing the various open access models have identified various roadblocks – some difficult and complicated and others much more tractable – that need to be addressed in the creation of open repositories[97]. These issues may be categorized according to scientific discipline, technological, institutional, management, economic, social, and legal areas. Only the latter three are addressed below[98].

5.1 Economic aspects

Cost. The question of cost, of course, is a threshold consideration for any activity. The costs of open repositories can range from a few thousand U.S. dollars for small operations based on a single server with little additional labor or technical resources, to $1 million or more for complex repositories that provide a lot of different functionalities[99]. Generally, cost need not be an unduly inhibiting factor. An important consideration with regard to an open institutional repository is that it almost invariably will build upon the institution's information infrastructure and management operations that already are being undertaken within the institution, and thus typically represents only an incremental additional expense. Most if not all of the technical elements are either off-the-shelf standard technologies or open-source software. An open institutional repository also provides a lot of value to the institution, as discussed above, and therefore can bring both tangible (in the form of more grants and other indirect economic benefits) and intangible (reputational, and other) returns that may well exceed the incremental costs. Most important in the context of this policy

[97] See, e.g., Swan, Alma, and Sheridan Brown (2005), "Open access self-archiving: An author study, Key Perspectives, Ltd., available at: http://eprints.ecs.soton.ac.uk/10999/; and Bjork, Bo-Christer (2004), "Open access to scientific publications – an analysis of the barriers to change", *Information Research*, 9(2).

[98] The institutional issues have already been discussed above.

[99] Another interesting cost comparison of 10 such repositories was compiled by Kemp, Rebecca at the University of North Carolina, available at: http://library.uncwil.edu/Faculty/kempr/listserv-summary-IR-open-source-costs.xls/.

debate – and also the most difficult to measure – are the costs of *not* providing open availability to this material, as discussed in Part II.

Open thematic repositories have a somewhat different cost structure. Although such repositories also will most likely be founded upon an existing information management operation at a major institution, they have different cost drivers based on their functionality. Unlike the open institutional repositories, whose principal sources are multidisciplinary and internal to the organization, the thematic repositories serve external authors in a specific research community. This means that the open thematic repositories may not have a broad internal organizational constituency, as in the case of an institutional repository. Because thematic repositories typically have more ambitious and complex operations, they also will tend to have higher operating costs than institutional ones.

Funding. The sources of funding form the other half of the financial equation. Because both types of repositories discussed here make their material openly available, they must have a funding model that generates revenue from sources other than the users. In all cases, the host institution will be directly subsidizing at least part of the costs of operation by virtue of providing some if not all of the facilities, technologies, and personnel. Thus the smaller-scale open institutional repositories can usually be funded as part of the host organization's overhead costs without the need for any (or substantial) additional funds from external sources. The various benefits derived from the operation of such a repository can be expected to offset the relatively modest operating costs, although this may be difficult to quantify in practice.

The more ambitious the scope of the repository the more will other funds need to be raised. In the case of a government repository – whether institutional or thematic – the funding must be authorized and appropriated annually as part of the institution's overall budget. In the case of a university of not-for-profit organization, the most common external funding sources are grants from government or

philanthropic organizations. There are not many other options[100]. Although one large grant or several smaller ones may be sufficient to launch a non-governmental repository and then keep it going as part of the institution's normal library or information management operations, a continuous source of grants is most likely necessary to maintain the largest repositories, particularly thematic ones.

Such potential funding uncertainties for the largest open repositories may raise a question of long-term sustainability. Fortunately, in practice such a worst-case scenario has not yet occurred, perhaps because such repositories have not been ubiquitous or long-lived yet, but more likely because just as the libraries that either preceded them or that they are a part of, their parent institution will not allow them to fail.

5.2 Social aspects

Author compliance. A major issue that has become quite evident in practice is a lack of compliance in submission of manuscripts by the authors of journal articles. As noted above, it is demonstrably in authors' interest to deposit their works, whether the repository is institutional or thematic. In practice, however, it has not been easy to achieve broad participation[101].

There are several reasons[102] that have been adduced to this, some that are common to both kinds of repositories and some that are unique to each type. A major reason appears to result from the authors' lack of knowledge of the repositories, the policies concerning them, or their potential benefits to the authors. Another is the perception by many authors that depositing the material is time consuming or onerous. A final major barrier is the authors' knowledge or perception that their publisher either does not allow or disapproves of such deposits, or that future acceptance of the authors' publications in peer-reviewed journals may be compromised.

[100] Institutional subsidies, volunteer work, and potential advertising revenues are additional possible sources of support. Bjork, *op. cit.*, note 97.

[101] Voluntary author deposits typically are at low levels, although they have been increasing. See Swan & Brown, *op. cit.*, note 97.

[102] *Ibid*. See also Bjork, *op. cit.*, note 97.

G. Sica (ed.) *Open Access: Open Problems*

More specifically, there are somewhat different dynamics associated with each type of repository that either motivates or dissuades authors from depositing their works. Institutional repositories can be more persuasive or prescriptive with their employees, who are under the direct control of their home institution. Thematic repositories have the potential to serve an author's interests better within a specific discipline collection and focus, but in practice this has not yet always proved to be persuasive.

There are several approaches that can be taken to improve upon this chronic lack of compliance. One is for the public research sponsors and repository staff to engage in more direct outreach to authors and educating them about the benefits of open availability, while debunking some of the extant misperceptions. This option has been relatively ineffective thus far, however, and also can be quite costly in its implementation because of its labor-intensive nature.

Another approach is for the research funders or the authors' institutional employers to either encourage or require the authors to retain copyright in their works, rather than assign copyright to the publishers. This option is discussed in more detail below. The benefit to authors in retaining copyright is that the authors will have greater legal certainty and perhaps greater flexibility in making their works openly available, thereby facilitating deposits to open repositories. Nevertheless, this option does not directly address the authors' reluctance or failure to deposit, but just helps overcome one of the most substantial hurdles to doing so.

The final approach is for the research funders or the institutional employers to either encourage or to require authors to submit a copy of their works to an open repository either immediately or within some prescribed period of time. Simply encouraging authors to deposit their works in open repositories thus far has been demonstrated to be ineffective, much like the outreach and education approach mentioned above. A requirement to do so would be much more effective, but has met with substantial resistance from the legacy publishers, especially if the requirement is for immediate deposit upon publication rather than following some embargo period of several months. Nevertheless, several universities and research

funders already require such deposits[103]. These policies are only now being fully debated and tested in different countries. If the goal of the research funders or institutional employers of the authors is to obtain the highest levels of open availability and return on research investment, however, the requirement for the authors to deposit a copy of their work immediately upon publication will be by far the most effective option.

Publisher compliance. As noted in Part III, many publishers have recently agreed to make all their restricted-access versions openly available following periods that typically span 6 or 12 months. Most publishers also either expressly allow, or do not object to, authors' requests to deposit a copy to their own institution's open repository. However, most publishers also resist any mandated deposits, particularly immediate ones. Additional points of contention involve what version should be deposited – the author's pre-publication version that has been accepted for publication, or the final published version – as well as who should deposit it, the author or the publisher. The issues associated with who retains copyright are discussed separately below.

The question of publisher compliance with policies or practices designed to increase open availability to the publicly-funded works that they publish is complex and has multiple levels. As the established intermediaries in the process of scientific communication, the publishers who operate on the legacy subscription-based business model have a vested interest in the status quo and perceive the greatest risk in any moves toward open availability. Commercial publishers clearly have the largest financial stake in this debate and are less closely aligned with the scientific communities than are the not-for-profit professional society publishers, whose authors and readers are primarily drawn from the ranks of their society memberships. For their part, the not-for-profit publishers may be

[103] Several universities and CERN require immediate self-archiving by their researchers. For the list and the policies see http://www.eprints.org/openaccess/policysignup/. The Wellcome Trust was one of the first major research funding organizations to require self archiving by its grantees. See: http://www.wellcome.ac.uk/doc_WTD002766.html.

G. Sica (ed.) *Open Access: Open Problems*

expected to be more responsive to the broader interests of the scientific community, while still concerned about the effects of greater openness on their revenues, as the "DC Principles" cited in Part III imply[104]. Finally, there are the emerging open access publishers, largely from the not-for-profits and from within the research community itself, who are demonstrating the feasibility of open access publishing as an alternative model. In any case, the business (and other) factors associated with the publishing of different types of journals in different disciplines can vary substantially[105].

While the publishers certainly may be expected to act in their own self-interest, they are subject to both direct and indirect influence by the various stakeholders in the research community. The most direct relationship is with the scientists themselves – the authors and the others who participate in the publishing enterprise, such as the editors and reviewers. These people have the greatest stake and interest on the supply side of the enterprise in how scientific publishing functions. They also have the most direct bargaining power to effect change. As has been noted repeatedly throughout this article, their

[104] The ambivalence and concerns of the not-for-profit publishers has been expressed in many places. See, e.g., The Royal Society of the United Kingdom (2005), The Royal Society of the United Kingdom, *op. cit.*, note 35; and Kaufman-Wills Group, LLC (2005), "The facts about Open Access", Association of Learned and Professional Society Publishers. While both the Royal Society statement and the ALPSP research report present a number of concerns about "open access" they both fail to adequately distinguish between open-access publishing and open repositories, with most of their concerns apparently focused on the former and not the latter.

[105] For a comprehensive list of the differences among disciplines in journal publishing, with particular focus on the relevant economic factors, see Peter Suber's compilation under "Lists Related to the Open Access Movement," last updated on November 6, 2005, available at: http://www.earlham.edu/~peters/fos/lists.htm/. A detailed examination of these differences is beyond the scope of this article, but a thorough analysis of these many factors would be invaluable in understanding the key pressure points and priorities of publishers, so that policies in support of open availability could be implemented in the most effective and appropriate way.

interest is (or should be) to maximize exposure to the fruits of their research.

The researchers also form the primary customer base – the demand for the outputs of the publishers – that provides another major regulating influence. The reading interests determine in large part what is used and not used, which journals succeed and which remain obscure or unprofitable. The university and institutional libraries are the other major players on the demand side. The libraries respond primarily to the interests expressed by their clients, and the publishers respond to the purchasing power, or the lack of it, by the libraries and other institutional subscribers. Scientists in developing countries also form an increasingly important block of user interests to whom the publishers and policy makers are now paying more attention. Finally, there also are the potential users in the broader society worldwide who form the silent majority of possible readers, but who are rarely considered or consulted in this context. All of these actual and putative users of the journal literature potentially want to get the broadest and least restricted access to it, and to promote their interests in this debate[106].

More powerful within the public research enterprise, but also much less directly connected to the publishing function, are the research institutions that employ the authors, the funding agencies that support the research on which the journal articles are based, and the government policy makers in the administrative and legislative branches of government. These institutional players have an interest in promoting the most efficient uses of the public investment in the research. They develop policies, regulate practices, and fund activities that are supposed to further the progress of science and the public interest.

[106] Interest groups have been formed for this purpose. In the United States, for example, such groups include the Scholarly Publishing and Academic Research Coalition (see http://www.arl.org/sparc) and its Open Access Working Group, and the Alliance for Taxpayer Access (see http://wwwtaxpayeraccess.org), which describes itself as "A diverse and growing alliance of organizations representing taxpayers, patients, physicians, researchers, and institutions that support open public access to taxpayer-funded research", primarily in biomedical research.

G. Sica (ed.) *Open Access: Open Problems*

The dynamics of these multiple interrelationships among the publishing intermediaries and the suppliers, users, regulators of the scientific literature are expressed and managed through their respective policies and practices. The manner in which each of these groups of stakeholders wields its influence and asserts its priorities, both collectively and individually, will determine how the open and restricted modes of access will evolve over time. The development of open repositories, particularly at the individual institutional level, is perhaps the least controversial method of mediating the respective interests of all the stakeholders worldwide in the transition from restricted to open availability of publicly-funded scientific information online.

5.3 Legal aspects

Copyright control is another potentially contentious issue that will require resolution over time. The question of who controls the copyright is not necessarily a barrier to open availability, but is the primary legal mechanism used to enforce the interests of the rights holder. As has already been pointed out, open availability can be provided with the express consent of the rights holder, whether or not full copyright is asserted.

Because researchers are the creators of the scientific literature, they have the original and moral proprietary interests in their works. In many cases, however, it is their employing institution that owns the copyright because it is considered a "work for hire" under their employment contract[107]. Whether it is the authors or their employers that are the owners of the copyright, they are the ones to decide

[107] In a recent informal survey, about one-third of U.S. universities asserted copyright over the works produced by their employees. However, many of the researchers employed by those institutions either do not realize that this is so or disregard it when dealing with their publishers. Presentation by Ginsburg, Jane (2004), summarized in *Electronic Scientific, Technical, and Medical Journal Publishing and Its Implications: Report of a Symposium*, National Academies Press, Washington, DC, pp. 41-43. As a general matter, the employing institutions have considerable actual or potential control over whether their employee author can assign exclusive copyright to the publisher.

whether to assign it to others. Historically, the entities to which copyright was most frequently assigned have been the publishers of the works[108]. However, there is no immutable requirement that this be so; indeed, most of the new open access publishers do not mandate it. Many of the publishers who operate on the legacy subscription model in fact will accept a non-exclusive assignment of copyright to publish the works. This practice itself already indicates that the ultimate control over the copyright is a negotiable matter and should be decided on its relative merits. For the transfer of the works to open repositories, both the private interest of the authors and the public interest are better served by having the author retain the copyright and having the flexibility of assigning non-exclusive distribution rights to the publisher of record, the repository, and all downstream users.

Unfortunately, there is a great deal of confusion by authors about why retention of copyright may be important, whether or not assignment of the their copyright bars them from depositing copies of their works in open repositories, and even whether they in fact own the copyright in the works that they produce. The clarification of this situation and the development of more effective and "open-access friendly" copyright licensing practices can be undertaken by the research funding and employing institutions in collaboration with the researcher authors and the publishers[109].

5.4 Concluding observations

The advance of digital networks and technologies provides unprecedented yet underexploited opportunities to promote the progress of science and society. The challenges in deriving maximum value from the technological opportunities are largely in changing the human systems – the institutional, legal, economic, and sociological aspects as discussed in this article – rather than in the technological aspects, which continue to improve even if the

[108] For a listing of existing "Publisher copyright policies & self archiving" compiled at the University of Nottingham, see http://www.sherpa.ac.uk/romeo.php/.

[109] The Research Councils of the United Kingdom are planning to do this, for example. See RCUK, *op. cit.*, note 11.

G. Sica (ed.) *Open Access: Open Problems*

human systems do not. To make progress on these human aspects, all of the stakeholders who are involved worldwide in public research and in the process of communicating research results must take part in the unfolding debate. These participants in the scientific enterprise should be engaged in this process at some level, because they have a vested interest in its outcome.

Up to this point, most of the advances toward openness to the information outputs from publicly-funded research have come from the bottom up, from many dedicated and visionary individuals and institutions. These actors have been the path makers in developing a broad range of initially disparate, but related institutional and policy initiatives in diverse information types, disciplines, and countries. As these projects proliferate and become well-established, they are forming a nascent, interoperable global information commons for public science[110].

Those who fund and regulate public science from the top down are beginning to take notice. They are starting to build upon the tactical successes of the path makers and integrating them into the broader national and international strategies for the investment and management of public science. A gradual restructuring of the scientific information sector and of the processes of scientific communication is thus now well underway, taking more complete advantage of the transformational capabilities of the digitally networked technologies. As part of this restructuring process, open repositories for the scientific literature, whether institutional or thematic, are coming to be recognized as a fundamental component of the public research infrastructure.

Paul F. Uhlir
Director, International Scientific and Technical Information Programs
The National Academies

[110] A Global Information Commons for Science Initiative is being launched by the international Committee on Data for Science and Technology (CODATA), in collaboration with many other international scientific and informatics umbrella organizations. For additional background see: http://www.codata.org/.

G. Sica (ed.) *Open Access: Open Problems*