

Le risorse digitali di matematica e fisica

Lucio Lubiana

Biblioteca SISSA – Scuola internazionale
superiore di studi avanzati, Trieste
lubiana@sissa.it

Una rassegna di fonti e strumenti

I fisici e i matematici sono stati i primi a creare risorse digitali nell'ambito della loro letteratura professionale, realizzando database di preprint e riviste elettroniche. Si tratta di fonti costituite dalla cosiddetta letteratura grigia scientifica non verificata, mentre quella controllata verrà successivamente pubblicata nelle riviste di fisica e matematica, che sono state le prime ad essere digitalizzate. Accanto a queste si vanno affermando nuove fonti digitali: gli archivi di e-print e i depositi istituzionali. Tutti questi strumenti purtroppo si sono sviluppati disordinatamente, senza linee prestabilite, mettendo così in seria difficoltà gli utenti e i gestori delle fonti informative.¹

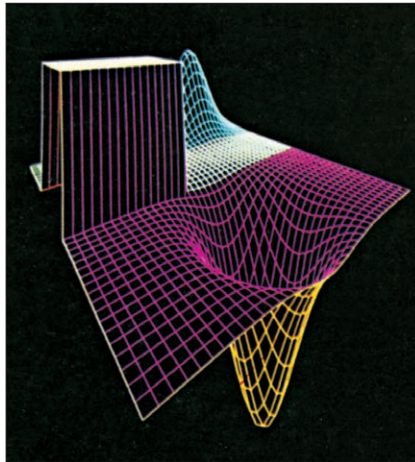
Gli ostacoli con cui ci si trova a dover fare i conti nell'archiviazione delle pubblicazioni scientifiche digitali sono rappresentati, secondo Richard R. Rowe, dalla mancanza di standard, dall'assenza di una consolidata tradizione, da instabili formati digitali, nonché dall'aumento enorme della letteratura digitale e dalla rigidità delle vigenti leggi sul copyright che continuano a limitare l'accesso legale alla letteratura scientifica archiviata.

Da molti anni sono questi i temi al centro del dibattito fra gli scienziati, gli editori e i bibliotecari.

I documenti digitali sono i più deteriorabili e "fragili" per quanto riguarda il mantenimento della loro integrità e risultano particolarmente difficili da conservare. Questo aspetto è cruciale per i matematici

e i fisici, in quanto la loro letteratura ha costantemente bisogno di essere ricercata, controllata e conservata. Inoltre, come sottolinea il matematico John Ewing, si rileva la mancanza di reali competenze nell'area dell'archiviazione digitale, dove non si è ancora sedimentata una grande esperienza.²

Per salvaguardare l'integrità e la "robustezza" delle informazioni digitali, si è così pensato di puntare sul-



l'archiviazione in più sedi. È questo il caso degli archivi di preprint di fisica e matematica, come l'archivio digitale PROLA (Physical Review Online Archive) dell'American Physical Society (APS), che prevedono l'archiviazione in "mirror" differenti per aumentarne l'accesso e per salvaguardarne l'integrità. Ma ciò può non bastare: l'identità di un documento può dipendere da file legati a sistemi più ampi (programmi, hardware, immagini), senza i

quali scompare l'autenticità della copia. In breve, sostiene ancora Ewing, "il contesto al quale il lavoro è legato è spesso essenziale per fare una copia fedele" e l'archiviazione deve essere in grado di riprodurre tale contesto. Si tratta della questione del "formato", che è oggi il problema centrale nel campo dell'archiviazione.³ Per risolverlo, alcuni recenti studi hanno proposto la migrazione continua dei dati e il cambiamento dei formati; ma cambiare formato non equivale a creare una nuova copia, in quanto richiede l'intervento dell'uomo che deve scegliere le informazioni da conservare (<http://www.iso.org/en/ISOonline.frontpage>).

Senza l'adozione di standard, però, diventa difficile leggere, valutare e conservare i documenti digitali. Quelli creati sul modello dello standard dell'Open Archives Initiative (OAI) possono essere più facilmente letti e conservati anche dagli scienziati dei paesi poveri.

A livello economico e organizzativo, come ha osservato Richard Rowe, si riscontrano tre modelli di archiviazione dell'informazione scientifica digitale:⁴

- 1) il modello tradizionale che fa capo alle case editrici, basato sul *peer review* e sui *referees*;
- 2) il modello rappresentato da BioMed Central e dalla Public Library of Science (PLOS) nel campo delle scienze bio-mediche, che accetta i principi dell'open access e assegna agli autori o alle istituzioni le spese di pubblicazione delle lo-

ro ricerche, mentre l'associazione non profit li rende accessibili liberamente. Questo modello è stato scelto, nel 2006, dal CERN di Ginevra per l'accesso libero e per il sostentamento delle riviste di fisica delle particelle elementari, ma non ancora per gli altri settori della fisica. L'obiettivo del movimento per l'open access, che ha il suo manifesto nella "Budapest Open Access Initiative" (2002), è quello di favorire l'autoarchiviazione (*self-archiving*) dei lavori scientifici in archivi digitali aperti e la creazione di una nuova generazione di riviste ad accesso libero basate sui principi della "Berlin Declaration on Open Access to Knowledge in Science and Humanities" (2003);

3) il modello incentrato sugli archivi digitali ad accesso libero in cui vengono depositati gli articoli scientifici per essere liberamente accessibili (es. gli archivi di preprints, ArXiv in fisica e matematica). In questo caso non c'è il filtro rappresentato dai *referees* che controllano la qualità degli articoli da pubblicare e da far circolare.

Gli archivi digitali sono di tre tipi: i depositi digitali istituzionali (per esempio di una università), gli archivi digitali disciplinari e i depositi digitali legati a specifici fondi. Vi è poi il processo di digitalizzazione delle vecchie annate di riviste di matematica che viene finanziato con fondi pubblici e i cui risultati possono essere, in gran parte, consultati liberamente online.

Le principali case editrici di ambito fisico-matematico (Elsevier, John Wiley & Sons, Springer Verlag) che dominano l'editoria scientifica (cartacea e digitale) sono molto preoccupate per l'impetuoso sviluppo dei modelli 2) e 3), che minacciano la loro sopravvivenza economica: anche per questo, nel 2006, hanno deciso di permettere agli autori di ripubblicare i loro lavori nei depositi digitali a libero accesso. Sul fronte pubblico, il National

Institute of Health statunitense, dal 2006, fa obbligo agli autori che hanno ricevuto fondi pubblici per le loro ricerche di renderli accessibili liberamente dodici mesi dopo la loro apparizione nelle riviste.

Il modello 1), sempre più "sotto assedio" da parte delle iniziative basate sull'accesso libero, viene considerato poco praticabile, in quanto l'archiviazione digitale comporta un onere finanziario non trascurabile e inoltre non c'è alcuna particolare garanzia di continuità delle stesse case editrici. Un'esperienza di questo tipo è stata comunque realizzata in Olanda, dove la casa editrice Elsevier, in collaborazione con la Biblioteca nazionale olandese, sta creando archivi digitali delle sue riviste e dei suoi preprint (è il progetto denominato e-Depot, nato nel 2002, a cui si sono aggiunti poi editori come Springer, Kluwer e Blackwell). Ancora Elsevier, nel 2005, ha raggiunto un accordo con Portico (www.portico.org) per depositarvi i suoi periodici elettronici. Nel complesso, però, i maggiori sviluppi stanno avvenendo nel campo degli archivi ad accesso libero. Un esempio tipico è rappresentato da DSpace (<http://www.dspace.org>), che è un archivio digitale istituzionale sviluppato congiuntamente dalle biblioteche del Massachusetts Institute of Technology e dalla Hewlett-Packard.⁵ Questo modello, basato sullo standard OAI e sullo standard Dublin Core per i metadati, è stato scelto da diverse università per la creazione delle loro collezioni digitali. Altri esempi di archivi digitali istituzionali sono gli e-print server (CogPrints, ArXiv, PubMed, BioMed Central, Astrophysical Data System). La loro proliferazione ha sollevato il problema dei diritti d'autore e quello relativo al sostentamento economico delle nuove iniziative editoriali.

Alcune società accademiche di fisica (AIP) permettono agli autori

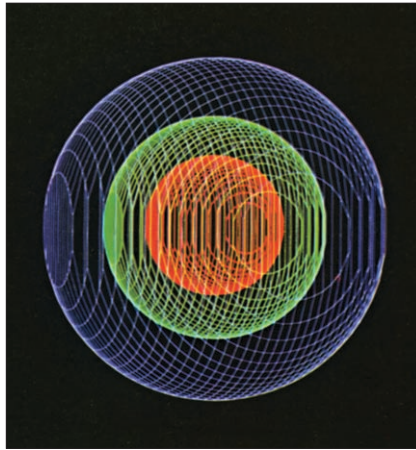
l'autoarchiviazione dei loro lavori scientifici in archivi digitali istituzionali. Alcune case editrici commerciali, invece, non lo consentono e si oppongono al decollo degli archivi "open access".

Gli archivi digitali disciplinari come ArXiv o CogPrint, scrive Antonella De Robbio, si "focalizzano su domini di soggetto specifico", mentre "i depositi istituzionali immagazzinano e rendono accessibile la produzione della propria istituzione".⁶ Fra le due forme di comunicazione, osserva sempre De Robbio, c'è un forte conflitto in quanto variano la qualità dei contributi archiviati e le forme di accesso e di circolazione, oltre che di archiviazione. In fisica e matematica le barriere fra i due tipi di archivi, l'archivio digitale di ArXiv e quelli delle riviste digitali o i database bibliografici, sono impercettibili.

Questa frantumazione dei confini fra i diversi contenitori di letteratura scientifica sta provocando anche un ripensamento dell'attuale sistema di valutazione dei lavori scientifici (*peer-review*) e della modalità di circolazione dei lavori. Si sta verificando una lenta erosione del monopolio esercitato dalle case editrici nella pubblicazione e nella disseminazione della letteratura scientifica e i fisici, in particolare, sono impegnati nella ricerca e nella proposta di nuove soluzioni efficaci.

Questi temi sono stati affrontati dall'International Congress of Scientific and Technical Information (<http://www.icsti.org/icsti.forum/33>), che ha proposto diversi modelli organizzativi per la creazione di un grande archivio digitale centralizzato per la scienza. Il dibattito che si è sviluppato ha portato all'elaborazione del progetto denominato "Digital Archives for Science and Engineering Resources" (DASER), che ha lo scopo di creare un deposito di informazioni scientifiche digitali fondato sullo stand-

ard OAI. Il progetto si propone, tra l'altro, di permettere lo scambio e la ricerca dei metadati standard fra differenti archivi digitali fra loro interoperabili. Attualmente sono stati censiti più di 779 archivi digitali che avevano implementato il modello OAI, ma il loro numero è in continua crescita. In base a recenti studi, circa il 43% degli editori permette l'auto-archiviazione di preprint e postprint negli archivi digitali istituzionali ("publisher copy-right policies: SHERPA, ROMEO"). In Italia PLEIADI (<http://www.openarchives.it/pleiadi>) è la piattaforma nazionale per la ricerca federata degli archivi digitali istituzionali. Gran parte di essi ha depositato e archiviato le informazioni in formato standard usando il modello OAI. Questo modello è appoggiato da numerose organizzazioni (università, biblioteche, case editrici), fra cui la Digital Library Federation e l'Online Computer Library Center (OCLC). Recentemente, però, un gruppo di scienziati riuniti dalla Microsoft, che hanno approvato la relazione *Towards 2020 sciences*, si sono pronunciati contro la creazione di un unico archivio digitale per la scienza. Un esempio di archiviazione digitale decentralizzata e federativa è costituito da CrossRef che è un'associazione non-profit di 78 società e case editrici che partecipano attivamente alla elaborazione di comuni standard di metadati da adattare ai vari archivi digitali (riviste, libri, letteratura grigia). Il modello aperto scelto da CrossRef è incentrato sul DOI (Digital Object Identifiers), un codice identificativo univoco e permanente che può essere incorporato nell'URL. La sua strategia organizzativa ed economica differisce da quelle centralizzate proposte da PubMed Central (PMC), della US National Library of Medicine, e da E-Biosci, sviluppato dalla European Molecular Biology Organization (EMBO).⁷



Alcuni pensano che i depositi digitali delle istituzioni (in particolare universitarie) potrebbero costituire dei modelli per la conservazione a lungo termine delle informazioni digitali, come nel caso dell'archivio digitale dei preprint di fisica e matematica ArXiv (<http://arXiv.org>), creato presso il Los Alamos National Laboratory, e ora conservato presso la Cornell University e da altri siti "mirror" presso altre istituzioni.⁸

Stevan Harnad è invece fautore della creazione di archivi di e-print aperti e interoperabili presso le istituzioni in cui lavorano gli scienziati, che potrebbero essere ricondotti a un archivio globale virtuale.

Ma l'accesso aperto alle risorse digitali non coincide con la creazione di dati aperti, informazioni digitali create secondo degli standard. Queste ultime sono necessarie per il web semantico che, in futuro, incrementerà lo scambio dei dati fra differenti archivi digitali. Attualmente manca un'infrastruttura in grado di gestire questi tipi di dati aperti. L'uso del formato XML favorirà, secondo Matt Cockerill, la conversione dei dati fra i diversi tipi di archivi digitali.⁹ Altri ritengono che sarà la diffusione del web semantico a incrementare la comunicazione, estendendo l'interoperabilità dei database digitali e fornendo nuovi strumenti per l'interazione con le collezioni multimediali.

Per raggiungere questi obiettivi il web semantico usa nuovi linguaggi basati sul RDF (Resource Description Framework), che è un metodo generale per rappresentare le proprietà delle risorse su web e le loro relazioni. Il web semantico viene applicato all'astronomia con il National Virtual Observatory e alla matematica, con il progetto inglese MONET, per rendere accessibili in formato web gli algoritmi e le formule matematiche con una varietà di pacchetti software.¹⁰

Forniamo ora una rassegna di alcune tra le principali fonti digitali di ambito matematico e fisico.

Fonti digitali per la matematica

Nel 2001 Antonella De Robbio ha presentato in modo esauriente le risorse online per la matematica in uno studio che a tutt'oggi costituisce il principale strumento di conoscenza per i bibliotecari.¹¹ Da allora però sono cambiati il numero e la qualità delle risorse elettroniche di matematica, che attualmente sono di tre tipi: le riviste elettroniche, i libri digitali e i database digitali bibliografici (MathSci Net e Zentralblatt MATH).

Per i matematici e i bibliotecari queste fonti hanno diverse strutture e funzioni, ma ultimamente (Convegno di Göttingen, 2003) si è auspicata la realizzazione di un accesso unico ai due principali database bibliografici di matematica e a quelli delle riviste digitali dello stesso settore. I matematici chiedono il superamento della separazione fisica e di ricerca fra le diverse fonti digitali, e la loro integrazione in comuni piattaforme informative. Ma questo, in molti casi, non è possibile, in quanto ogni casa editrice usa diverse politiche di accesso alle proprie fonti.

Un'importante proposta di integrazione fra fonti digitali diverse in matematica è rappresentata dal

progetto “The Digital Mathematics Library” (DML), che ha l’obiettivo di creare un database digitale centrale di tutta la letteratura matematica mondiale prodotta, accessibile a chiunque in possesso di un computer collegato tramite Internet. L’ideatore di questo ambizioso progetto è il matematico americano John Ewing, direttore esecutivo dell’American Mathematical Society. La grande visione del DML è quella di coordinare i diversi progetti di retro-digitalizzazione della letteratura di matematica con l’adozione di comuni standard tecnici capaci di facilitare i vari tipi di legami fra le fonti digitali.¹²

Un gruppo di matematici europei ha proposto nell’aprile del 2003 alla Commissione europea un progetto di finanziamento quinquennale per la realizzazione di una Digital Mathematics Library in Europa (DML-EU), che porti alla digitalizzazione delle riviste cartacee di matematica di ogni paese europeo. Questo progetto ricalca sostanzialmente quello americano, soltanto che ne delega la realizzazione alle singole nazioni.¹³

Fra le principali difficoltà riscontrate nella realizzazione della DML americana ci sono quelle relative alla negoziazione dei diritti d’autore, alla scelta dei libri e delle riviste da inserire nel progetto di digitalizzazione e quelle riguardanti i formati di memorizzazione e di archiviazione che possano essere accessibili a lungo termine. Questi nodi, ancora irrisolti, potranno essere superati con l’adozione dell’odierna tecnologia e anche perché la matematica è una disciplina ideale per la retro-digitalizzazione. Il Committee for Electronic Information and Communication (CEIC) dell’Unione matematica internazionale (IMU) nell’aprile del 2004 ha approvato una risoluzione per l’informazione e la comunicazione, in cui si raccomandava la pubblicazione in formati digitali stand-

ard. Si accennava alla necessità di creare una netta distinzione fra la pubblicazione e la validazione dei lavori matematici. Il comitato raccomandava, inoltre, l’archiviazione della letteratura digitale di matematica sotto il controllo della comunità accademica, per garantirne qualità e integrità (<http://www.ceic.math.ca/publications/racomendations>). Seguendo questi suggerimenti diverse società accademiche, biblioteche, case editrici hanno iniziato a digitalizzare la passata letteratura di matematica. I principali archivi digitali di matematica (la lista è incompleta, perché alcuni sono progetti) sono:

– JSTOR (Journal STORage, <<http://www.jstor.org>>), un archivio digitale interdisciplinare in cui sono, per ora, presenti diciassette riviste di matematica pubblicate dalla American Mathematical Society e da altre società (ma l’archivio si sta estendendo);

– DIEPER (Digitised European Periodicals, <<http://dieper.aib.unilinz.ac.at>>), digitalizzazione della rivista “Monatshefte für mathematik und physik” (1890-1918);

– Biblioteka Wirtualna Matematyki (<http://matwbn.icm.edu.pl>), che comprende le riviste “Fundamenta mathematicae” (1920-1993), “Studia mathematica” (1929-1964), “Prace matematyczne-fizyczne” (1888-1952) e alcuni libri elettronici di matematici polacchi;

– “Revista de Matematicas Aplicadas” (1994-2002) del Departamento de Ingeniería matemática Universidad de Chile (<http://www.dim.uchile.cl/revmat.html>);

– EMIS (European Mathematical Information Service, <<http://www.emis.de>>) e ERAM (Electronic Research Archiv for Mathematics), progetti di digitalizzazione della rivista “Jahrbuch über die Fortschritte der Mathematik” (1868-1942), con la creazione di un archivio digitale e con legami a libri digitali del progetto “Digital Library” della

Cornell University e del progetto “Gallica” (quest’ultimo sviluppato dalla Bibliothèque nationale de France, <<http://gallica.bnf.fr>> e dalle riviste “Journal de Mathématiques Pure et Appliquées”, 1936-1880, e “Comptes Rendus Hebdomadaires des Séances de l’Académie des Sciences”, 1835-1930);

– NUMDAM (Numérisation de Documents Anciens Mathématiques, <<http://www.numdam.org>>), che ha tra l’altro digitalizzato i seguenti periodici: “Annales de l’Institut Fourier” (1949-1997), “Bulletin de la Société Mathématique de France” (1872-1997), “Annales Scientifiques de l’Ecole Normale Supérieure” (1864-1997), “Journées Equations aux derives partielles” (1974-2000), “Mémoires de la Société Mathématique de France” (1964-1992), “Publications Mathématiques de l’IHES” (1959-1997);

– il progetto Euclid (<http://ProjectEuclid.org/mmj>), che digitalizza la rivista “Michigan Mathematical Journal” dal 1952 a oggi;

– SciELO (Scientific Electronic Library Online, <<http://www.scielo.cl/scielo.php>>), che digitalizza la rivista “Proyecciones. Revista de Mathematica” (2000-2002);

– WDML-Göttingen (Digitalisierungszentrum, <<http://www.sub.uni-goettingen.de/gdz>>, World Digital Mathematical Library): nell’archivio sono presenti le riviste di matematica pubblicate dalla casa editrice Springer e da altre case editrici.¹⁴

È stato creato un legame elettronico fra le citazioni del principale archivio bibliografico di matematica, MathSciNet, e i documenti elettronici conservati nell’archivio JSTOR. Un’operazione simile è stata fatta anche fra le citazioni bibliografiche dell’archivio bibliografico Zentralblatt MATH e gli articoli conservati nell’archivio del progetto WDML-Göttingen.

Per i libri digitali di matematica è stato realizzato un catalogo centra-

lizzato presso l'Università Joseph Fourier di Grenoble ("The Cellule MathDoc") che censisce i libri di quattro siti web: il "Digital Math Books", della biblioteca della Cornell University, i libri di matematica della collezione "Gallica" della Bibliothèque nationale de France, i libri digitali del progetto "WDML-Göttingen", presso il "Göttinger Digitalisierungszentrum", e quelli della collezione storica dell'Università del Michigan (<http://math-doc.ujf-grenoble.fr/LiNuM/>).

Nel campo dei database bibliografici i due principali archivi bibliografici di matematica sono: Math SciNet e Zentralblatt MATH. Math SciNet (<http://www.ams.org/msn.html>) permette l'accesso alle citazioni bibliografiche raccolte dalle riviste "Mathematical Reviews" e "Current Mathematical Publication" dalla loro nascita a oggi. Da questo database si possono raggiungere, dal 2002 anche tramite l'openURL, diciassette riviste di matematica depositate nell'archivio digitale JSTOR, le riviste elettroniche delle case editrici Springer, Birkhauser, Elsevier. Il Zentralblatt MATH, invece, è un database che contiene oltre due milioni di citazioni bibliografiche estratte da più di 2.300 riviste, e copre il periodo dal 1868 a oggi. Anche questo archivio bibliografico ha realizzato diversi siti "mirror", fra cui uno in Italia (Lecce, SIBA): Spagna, Francia, USA (Berkeley e Ithaca), Brasile, Canada, Cina e Germania (Göttingen e Berlino). Recentemente, col progetto LIMES, finanziato dalla Commissione europea e dall'Istituto di fisica FIZ di Karlsruhe, si cerca di estendere a livello europeo questo database bibliografico digitale di matematica. Recentemente, a un convegno che si è tenuto presso l'Università di Göttingen, un ristretto gruppo del "Mathematical Reviews", del "Zentralblatt für Mathematik und ihre Grenzgebiete" e dei rappresentanti dei progetti di

digitalizzazione di matematica ha proposto di sviluppare standard tecnici per facilitare il legame elettronico fra i database bibliografici di matematica (Math SciNet e Zentralblatt MATH) e gli archivi digitali delle riviste di matematica.

Fonti digitali per la fisica

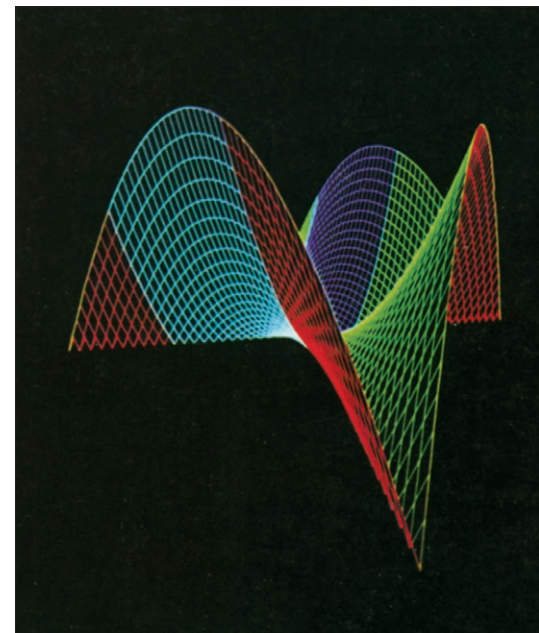
Storicamente, la creazione dei primi archivi elettronici e delle prime riviste elettroniche è avvenuta nei dipartimenti universitari di fisica più che di matematica. Sono stati i fisici, pertanto, i più sensibili interpreti dell'esigenza di creare archivi digitali aperti, accessibili, costruiti con standard condivisi.

Non è un caso, dunque, che sia stata la Società europea di fisica, nel suo congresso del 2000, ad approvare una risoluzione di richiesta di standard condivisi nella creazione e conservazione dei documenti digitali della materia. Questa esigenza non era presente nel 1991, quando lo scienziato Paul Ginsparg, presso il Laboratorio di fisica delle alte energie di Los Alamos (New Mexico), creò il primo archivio digitale di preprint di fisica, matematica e astrofisica, che divenne ben presto uno dei principali strumenti di circolazione dei risultati dell'attività di ricerca fra i fisici. Attualmente questo archivio di letteratura grigia è ospitato presso la biblioteca della Cornell University¹⁵ e riceve mensilmente più di 3.000 nuovi articoli, ma essi non sono controllati per la validità delle informazioni scientifiche che contengono. Un altro database bibliografico di letteratura grigia usato dai fisici è Slac SPIRES. Per molti fisici la pubblicazione dei loro lavori nelle riviste cosiddette "peer-reviewed" costituisce un elemento essenziale di valutazione, anche in funzione della loro progressione di carriera in ambito accademico. Questo modello di au-

toarchiviazione digitale ha oscurato i confini fra la letteratura grigia (preprint) e quella formale pubblicata nelle riviste. Non tutte le case editrici, però, permettono l'accesso libero e l'autoarchiviazione, prima o dopo la pubblicazione, degli studi degli scienziati.

La rivista "Science", ad esempio, non permette l'autoarchiviazione degli articoli prima della loro pubblicazione. Altre case editrici, come l'American Physical Society, la consentono in appositi archivi di preprint. Questo nodo irrisolto costituisce un elemento di impedimento alla libera e sicura circolazione dei lavori scientifici.

Nuovi sistemi di filtraggio dei risultati della ricerca dovranno essere elaborati per separare la validazione dalla pubblicazione. Per i fisici questo confine è molto labile, in quanto gran parte di quello che pubblicano nell'archivio di preprint viene poi validato con la pubblicazione nelle riviste scientifiche. Il fisico americano Jorge Hirsch ha proposto un nuovo modello di validazione dei lavori scientifici (*b-index*) basato su un indice derivato dal numero di volte che l'articolo viene citato e pren-



dendo in considerazione anche la qualità dei lavori pubblicati.

Questi filtri di selezione mancano negli archivi digitali di preprint, ma per motivi di sicurezza e di accesso sono state create delle copie digitali in diverse sedi: in Italia, per esempio, il “mirror” degli archivi elettronici della Cornell University si trova nel sito Internet della SISSA di Trieste (<http://bab.bage.sissa.it>). Analoga iniziativa è stata presa in Francia, nel 2000, dal CNRS (Centre National de la Recherche Scientifique).¹⁶

Nel 2002 l'archivio digitale dei preprint di fisica ArXiv è stato testato, con altri depositi digitali, per provare la validità della versione 2.0 dello standard OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting). Quest'ultimo standard è stato adottato (maggio 2002) da 77 archivi digitali, non solo di fisica, e il loro numero sta continuamente crescendo. Il protocollo OAI è stato adottato anche dal CERN Document server software per la ricerca di 550.000 record di metadati di documenti digitali prodotti dai fisici del CERN di Ginevra.¹⁷

Un'altra applicazione di questo protocollo si ha con il portale Torii (<http://torii.sissa.it>), realizzato dalla SISSA di Trieste con fondi europei all'interno del progetto TIPS, che permette l'accesso integrato e profondamente innovativo agli archivi digitali aperti (per ora gli archivi di preprint Los Alamos ArXiv, preprint di scienze cognitive, Biomed). Il portale Torii usa il protocollo standard dell'OAI come strumento base di comunicazione fra il portale e i sottostanti servizi (archivi, servizi, riviste elettroniche), operando in modo trasparente.¹⁸

Un altro importante strumento bibliografico di letteratura grigia di fisica delle particelle elementari è SPIRES (Stanford Public Information Retrieval System), realizzato presso l'Università di Stanford, che permet-

te l'accesso a oltre 500.000 articoli di fisica, compresi articoli di riviste, preprint, rapporti tecnici, atti di congressi e tesi, indicizzati dalla Biblioteca universitaria di Stanford e da DESY, a partire dal 1974. Di questo archivio sono stati costruiti siti “mirror” in Gran Bretagna, Giappone, Russia e Germania.¹⁹

Ma tutte queste fonti digitali di fisica sono disomogenee e per tale motivo alcuni studiosi hanno proposto l'interoperabilità fra di esse, gli e-print, gli archivi digitali distribuiti e le riviste elettroniche. Lo strumento che lo permette è il protocollo standard OAI-PMH, adottato da diversi archivi digitali, tra cui quello di astrofisica ADS (NASA Astrophysics Data System) e di preprint di fisica della Cornell University. Ma non tutte le risorse digitali sono state costruite con questo standard.²⁰

Gli astronomi hanno creato una rete di archivi digitali distribuiti in rete, incentrati sul database bibliografico ADS (<http://adswwww.harvard.edu/>), e hanno sviluppato standard disciplinari e strumenti di interoperabilità come, ad esempio, il FITS che memorizza in formato standard i dati delle immagini, degli spettri e le tabelle. ADS gestisce quattro database bibliografici diversi, con più di due milioni di record di astronomia, strumenti, fisica, geofisica e preprint di astrofisica. Esso, inoltre, dà l'accesso a diverse risorse digitali esterne, come articoli di riviste a testo completo, cataloghi e archivi.

Un altro strumento di ricerca e di identificazione usato dagli astrofisici è il “bibcode” (la descrizione dei riferimenti bibliografici con diciannove caratteri), adottato da ADS, dalle riviste elettroniche, dagli archivi degli osservatori, e che costituisce un legame preciso agli articoli pubblicati. Si pensa di estendere la sua applicabilità ad altri sistemi bibliografici digitali in rete, e di renderlo interoperabile

con lo standard DOI. Un altro esempio di uso del web semantico in astrofisica è il legame realizzato fra le diverse risorse di astrofisica e il progetto inglese Monet (<http://monet.nag.co.uk/cocoon/monet>).

Ma la principale iniziativa bibliografica proposta dagli astrofisici, nel 2002, è stata quella di fondare l'International Virtual Observatory (IVO), con l'intento di integrare fra loro le principali basi di dati di astrofisica, usando tecnologie, dati e software standard.

Nel campo della fisica, invece, uno dei più importanti archivi digitali delle riviste di fisica è il PROLA (Physical Review Online Archive). Al suo fianco sono stati creati gli archivi digitali delle riviste elettroniche delle principali case editrici scientifiche (Springer, IOP, European Physical Society, Elsevier, World Scientific, American Institute of Physics). Non è ancora chiaro se questi archivi verranno gestiti direttamente dagli editori, oppure dati in gestione a biblioteche o consorzi. Negli Stati Uniti gran parte delle case editrici accademiche ha assegnato la gestione dei propri archivi digitali di riviste a società indipendenti, come la Highwire Press, mentre soltanto alcune li gestiscono in proprio. Il PROLA è il principale strumento bibliografico per l'accesso alle riviste elettroniche diffuse dall'American Physical Society (APS), che pubblica “Physical Review”, “Physical Review Letters”, “Reviews of Modern Physics”. L'archivio digitale completo è costituito da “Physical Review” (dal 1893, con le varie sezioni), “Physical Review Letters” (dal 1958) e “Reviews of Modern Physics” (dal 1929). Una copia di questo archivio è depositata presso la Library of Congress e un'altra presso la Biblioteca della Cornell University (<http://cornell.mirror.aps.org/info>). La maggior parte della collezione è costituita da immagini scannerizzate, in formato

immagine GIF o in formato PDF, delle riviste a stampa. Le informazioni bibliografiche sono in formato SGML. PROLA ha creato il suo database bibliografico digitale con il linguaggio XML con cui sono stati costruiti l'indice di ricerca e i legami di riferimento ai documenti originali. Nell'ottobre del 2003 il Los Alamos National Laboratory ha ottenuto la licenza dall'APS di conservare l'intera collezione digitale delle sue riviste e l'archivio PROLA. L'APS e il centro di Los Alamos usano i protocolli OAI per mantenere la copia sincronizzata con la versione originale dell'archivio.

I due enti si impegnano, inoltre, a usare un modello standard di distribuzione del contenuto degli archivi usando l'OAI-PMH. Il modello del preprint di fisica di Los Alamos è stato poi imitato anche dalle scienze cognitive, e ultimamente anche dalla biologia. In quest'ultimo caso sono state sollevate, da diversi studiosi, fondate preoccupazioni sulla accessibilità delle informazioni mediche non controllate da parte di medici e pazienti.²¹

Accanto a questo modello di autoarchiviazione libera e non controllata dei risultati della ricerca, i fisici hanno creato le loro prime riviste completamente elettroniche. Esse costituiscono la fonte principale per l'archiviazione digitale a lungo termine della letteratura controllata di settore. Questi temi sono stati al centro del convegno che si è tenuto nel settembre del 2005 presso il CERN di Ginevra, al quale sono intervenuti fisici, bibliotecari e case editrici. Ciascuno di essi ha analizzato, dal proprio punto di vista, l'impatto che l'editoria elettronica ha nella fisica (<http://open-access.web.cern.ch/Open-Access/20050916.html>). Da questi dibattiti è scaturita l'iniziativa del CERN, intrapresa nel 2006 e appoggiata da diverse società accademiche, di costituire un fondo

per finanziare nuove iniziative editoriali digitali di fisica delle particelle elementari.

Sono stati i fisici della SISSA di Trieste a creare, nel 1998, una delle principali riviste di fisica delle particelle elementari, chiamata JHEP ("Journal of High Energy Physics", <<http://www.iop.org/journals/jhep>> o <<http://jhep.sissa.it>>). Di questa rivista elettronica si sono creati, per motivi di sicurezza e di integrità delle informazioni digitali, due archivi digitali: uno presso la casa editrice IOP e l'altro, di riserva, ancora presso la SISSA. A questo modello si sono ispirate, dal 2002 al 2005, le riviste elettroniche JCAP ("Journal of Cosmology and Astroparticle Physics", <<http://www.iop.org/journals/jcap>>), JCOM ("Journal of Science Communication", <<http://jcom.sissa.it>>) e POS ("Proceedings of Science", <<http://pos.sissa.it/>>). Il modello JHEP ha sollevato, quasi subito, il problema dei nuovi metodi di pubblicazione, quello dei costi delle pubblicazioni delle riviste digitali e quello della loro conservazione.

Molte case editrici commerciali si oppongono a questo modello editoriale, in quanto lederebbe i loro interessi commerciali.

Per la conservazione a lungo termine di queste riviste digitali gran parte di esse viene memorizzata nel formato previsto dall'OAI (<http://www.openarchives.org>) e nel formato PDF.

La casa editrice di letteratura scientifica Elsevier ha invece deciso di conservare la sua letteratura digitale presso la Biblioteca nazionale olandese. Altre case editrici si sono accordate con biblioteche o enti terzi per la conservazione della letteratura scientifica digitale.

Questi nuovi metodi di pubblicazione e conservazione delle risorse digitali hanno, per ora, sollevato un ampio dibattito tra i fisici, fra i fautori degli *open archives*, gli ar-

chivi di preprint, e quelli delle riviste fondate sul metodo della *peer review*. Comunque con il modello JHEP – scrive Bonora, uno degli ideatori della rivista – “il mondo scientifico riprende il controllo non solo degli aspetti scientifici, ma anche di quelli economici delle riviste specializzate”.²²

Diverso è il discorso dei libri digitali di fisica che si stanno propagando lentamente fra i fisici, essendo la loro diffusione ostacolata dai formati di codifica, dai programmi di lettura usati e dai dispositivi hardware usati.

Nel campo delle riviste si stanno invece affermando le seguenti iniziative editoriali: Ebrary, NetLibrary e Knovel. Si tratta di piattaforme informatiche che usano strumenti di consultazione e di diritti d'autore diversi, e anche i titoli contenuti sono differenti.

La gestione delle informazioni digitali è eterogenea fra questi strumenti commerciali e quelli proposti da GooglePrint (<http://www.books.com>), Amazon, dal progetto OCA (Open Content Alliance) della Microsoft e Yahoo e dal progetto “Million Book Project” della National Science Foundation della Carnegie Mellon University di Pittsburgh.

I due modelli economici che si stanno imponendo sono: uno incentrato sugli abbonamenti a un pacchetto di libri e l'altro su acquisti individuali. In entrambi i casi i contenuti sono separati dalle forme di accesso e di archiviazione.

Conclusioni

Non è ancora molto chiaro se le fonti digitali di matematica e di fisica verranno conservate dalle biblioteche o dalle case editrici, in quanto non è stato ancora trovato un nuovo modello economico che dovrebbe supportare questa finalità. Si stanno facendo alcuni tenta-

tivi – è il caso di PLOS e Biomed Central – di passare dagli abbonamenti tradizionali, effettuati tramite le biblioteche, al modello ad accesso libero, incentrato sul pagamento delle spese di pubblicazione delle risorse scientifiche da parte degli scienziati e delle loro istituzioni di ricerca. L'AIP ha pianificato di pubblicare, dal 2007, diversi titoli di periodici ad accesso libero.

Questo nuovo modello potrebbe sovvertire le basi economiche dell'archiviazione delle risorse digitali che passerebbero dalle biblioteche alle case editrici, o a enti terzi.

C'è inoltre la richiesta degli scienziati di integrare fra loro le diverse fonti digitali di matematica e fisica, per facilitarne la circolazione e l'accesso. In particolare essi desiderano l'integrazione fra i database bibliografici, le riviste e i libri digitali, utilizzando sofisticati metodi di autenticazione e di rispetto dei diritti d'autore. Ma questo obiettivo è ancora molto lontano e il suo raggiungimento è ostacolato da molti vincoli e impedimenti.

L'ultima esigenza manifestata dagli scienziati è quella di creare sofisticati strumenti di ricerca, come il *semantic web* di Tim Berners-Lee o il *data mining*, per migliorare l'accesso alla letteratura scientifica. A questa esigenza viene incontro lo sviluppo di nuovi software capaci di esplorare i database aperti, come ad esempio il Pub Med database o il progetto "Bio Text" della Università di Berkeley, California. Tali progetti usano gli attuali standard e molti scienziati auspicano la loro adozione anche da parte delle case editrici, per permettere l'accessibilità di tutta la letteratura scientifica archiviata in formato digitale. Ma un'effettiva integrazione informativa esige un accordo su una comune struttura semantica dei dati (e recentemente sono aumentati i dubbi sulla reale realizzabilità di un unico modello onto-

logico globale), mentre le case editrici si oppongono a questo processo, in quanto temono di perdere il monopolio della produzione e diffusione della letteratura scientifica.²³

Note

¹ Cfr. SEAMUS ROSS, *Position paper on integrity and authenticity of digital cultural heritage objects*, round table held in Barcelona on May 6th, 2002, "Digi CULT", p. 7, e ANN J. WOLPERT, *The future of electronic data*, "Nature", (2002), 420, p. 17.

² Cfr. JOHN EWING, *Twenty centuries of mathematics: digitizing and disseminating the past mathematical literature*, "Notices of the AMS", 49 (2002), 7, p. 775.

³ *Ivi*, p. 776.

⁴ Si vedano a questo proposito i seguenti contributi, comparsi sul sito di "Nature": RICHARD R. ROWE, *Digital archives*, <<http://www.nature.com/nature/debates/e-access/Articles/rowe.html>>; MATT COCKERILL, *Distributed and centralized technologies*, <<http://www.nature.com/nature/debates/e-access/Articles/cockerill.html>>.

⁵ Cfr. *Open access now. A journey into Dspace. Biomedcentral*, <<http://www.biomedcentral.com/openaccess/archive/>>, e KOROLYN HANK, *Digital curation and trusted repositories, seeking success*, JCDL 2006 Workshop Report, "D-Lib Magazine", 12 (2006), 7/8, <www.dlib.org/dlib/july06/hank/07_hank.html>.

⁶ ANTONELLA DE ROBBIO, *Auto-archiviazione per la ricerca: problemi aperti e futuri sviluppi*, Convegno "Comunicazione scientifica ed editoria elettronica: la parola agli autori", Università degli studi di Milano, 20 maggio 2003, p. 3. Si veda anche ANN J. WOLPERT, *The future of electronic data*, cit., p. 17-18; *Quality and value: models of quality control for scientific research*, "Nature", 2006, <<http://www.nature.com/nature/peerreview/debate/nature05031.html>>; *Systems: online frontiers of the peer-reviewed literature*, "Nature", 2006, <<http://www.nature.com/nature/peerreview/debate/nature05030.html>>.

⁷ DECLAN BUTLER, *EMBO backs single electronic repository*, "Nature", (1999), 400, p. 97, <<http://www.nature.com/doi/10.1038/21948>>; ID., *Scientific*

publishing: who will pay for open access, "Nature", (2003), 425, p. 554-555, <<http://www.nature.com/nature/journal/v425/n6958/full/425554a.html>>.

⁸ Cfr. RAYM CROW, *The case for institutional repositories: a SPARC position paper*, <<http://www.arl.org/sparc/IR/ir.html>>.

⁹ Cfr. MATT COCKERILL, *Distributed and centralized technologies: complementary tools to build a permanent digital archive*, <<http://www.nature.com/nature/debates/e-access/Articles/cockerill.html>>, e STEVAN HARNAD, *Commentary*, "Nature", (2001), 410, p. 1024-1025; EDWARD M. CORRADO, *The importance of open access, open source, and open standards for libraries*, "Issues in Science and Technology Librarianship", (2005), <<http://www.isl.org/05-spring/article2.html>>.

¹⁰ Cfr. A. SZALAY – J. GRAY, *The worldwide telescope*, "Science", 14 September 2001, <<http://www.sciencemag.org/cgi/content/abstract/293/5537/2037>>; JAMES HENDLER, *Enhanced: science and the semantic web*, "Science", 299 (2003), 5606, p. 520.

¹¹ ANTONELLA DE ROBBIO, *Le risorse online per la matematica nel virtual reference desk scientifico*, "Bibliotime", 4 (2001), 1, <<http://didattica.spbo.unibo.it/bibliotime/num-iv-1/derobbio.htm>>.

¹² JOHN H. EWING, *Twenty centuries of mathematics: digitizing and disseminating the past mathematical literature*, "Notices of the AMS", 49 (2002), 7, p. 771-777.

¹³ Per questo progetto, cfr. ANTONELLA DE ROBBIO, *Digital Math Library*, "Biblioteche oggi", 21 (2003), 6, p. 47-56.

¹⁴ ALLYN JACKSON, *The Digital Mathematics Library*, "Notices of the AMS", 50 (2003), 8, p. 920.

¹⁵ STEVAN HARNAD, *Correspondence*, "Nature", (1999), 401, p. 423.

¹⁶ Cfr. SUSAN Y. CRAWFORD – JULIE HURD – ANN C. WELLER, *From print to electronic: the transformation of scientific communication*, Medford (NJ), Information Today, 1996, p. 69-75, e DECLAN BUTLER, *French take physics archives into the future*, "Nature", (2000), 407, p. 825; *Number theory*, "Physicsweb", 16 August 2005 (<<http://physicsweb.org/articles/nws/9/8/1>>); BAIL PHILIP, *Index aims for fair ranking of scientists*, "Nature", (2005), 436, p. 900.

¹⁷ MICHAEL NELSON, *New developments*

in OAI (OAI-PMH2), proceedings 1st. "Open Archives Forum Workshop", Pisa, 14 August 2002; MARTIN VESELY, *CERN document server software*.

¹⁸ Cfr. SARA BERTOCCO, *Torii, an open portal over open archives*, "High Energy Physics Libraries Webzine", (2001), 4, <<http://library.cern.ch/HEPLW/4/papers/4/>>.

¹⁹ Cfr. SUSAN Y. CRAWFORD – JULIE HURD – ANN C. WELLER, *From print to electronic*, cit., p. 69-70.

²⁰ FRANÇOISE GENOVA, *Online information in astronomy: from networking to a virtual observatory*, delivered at the 1st "Open Archives Forum Workshop", Pisa, 14 August 2002; DECLAN BUTLER, *Astronomers give virtual observatory a real future*, "Nature", (2002), 417, p. 777.

²¹ Cfr. ID., *Biologists join physics preprint club*, "Nature", (2003), 425, p. 548.

²² LORIANO BONORA, *L'evoluzione dell'editoria scientifica e il modello JHEP*, Convegno "Editoria elettronica, open archives e comunità digitale", Trieste, 13 maggio 2004, <<http://www.sissa.it/library/convegno.html>>.

²³ Cfr. *Machine readability*, "Nature", (2006), 440, p. 1090, e JOHN E. ENDERBY,

Considering multiple flavors, "Science", 14 April 2006, p. 200; ANDREAS VON BUBNOFF, *The real death of print*, "Nature", (2005), 438, p. 550-552; GINO RONCAGLIA – F. MESCHINI, *E-book per gli studenti*, in *E-book: risorse attuali e*

prospettive future: seminario con tavola rotonda organizzato da Cenfor International, Milano, Bibliostar 2005, Palazzo delle Stelline, 18 marzo 2005, a cura di Serena Sangiorgi e Anna Merlo, Roma, AIDA, 2006, p. 23-39.

Abstract

This article describes the creation of different digital sources in mathematics and physics. These sources were journals, bibliographic databases, digital books and digital repositories. At the end of the twentieth century, the physicists created, at Los Alamos Laboratory, the digital archive of preprint (arXiv). After that period were formed the first digital journals and digital repositories. Great part of the physics and mathematics journals in print were digitised. Their proliferation has raised the problem of their standardization of formats and data.

For this reason was designed the Open Archives Initiative Protocol (OAI) and the first metadata standard for the description of digital objects, Dublin Core.

New tools for identification of digital objects were prepared like DOI (Digital Object Identifiers), OpenURL framework for context-sensitive linking and XML language and others.

All of them have made easier the description, access (copyright) and the circulation of scientific literature.