

Gli archivi digitali di matematica e fisica

Lucio Lubiana
SISSA

Introduzione

La sfida nella archiviazione delle pubblicazioni scientifiche digitali è dovuta, scrive Richard R. Rowe, alla mancanza di standard, all' assenza di stabili tradizioni in quanto i formati digitali sono instabili, all' aumento enorme della letteratura digitale prodotta e alla rigidità delle correnti leggi del copyright che continuano a limitare l' accesso legale alla letteratura archiviata.

I documenti digitali sono, per gli archivisti, i più deteriorabili e più fragili nel mantenimento della loro integrità e sono, anche, difficili da conservare.¹ Accanto a questi problemi c' è quello della mancanza, scrive il matematico John Ewing, di reali esperti nell' area dell' archiviazione digitale dove non c' è ancora una grande esperienza in questo campo.²

Per mantenere la loro integrità e la loro "robustezza", scrive J. Ewing, le copie digitali sono state archiviate in uno o due posti differenti.

E' questo il caso degli archivi di preprint di fisica e matematica, l' archivio digitale PROLA ("Physical Review Online Archive") dell' "American Physical Society" (APS) che sono stati archiviati in "mirror" differenti per aumentarne l' accesso e per mantenere la loro integrità.

Ma questa non è detto che rimanga la stessa nel corso del tempo e questo è dovuto ai file legati a più ampi sistemi (programmi, hardware, immagini) senza i quali scompare l' autenticità della copia.

In breve, scrive J. Ewing, "il contesto nel quale il lavoro è legato è spesso essenziale per fare una copia fedele" e l' archiviazione deve essere in grado a riprodurre questo contesto.

Questo rappresenta, per J. Ewing, il problema del "formato" che è il problema centrale dell' archiviazione.³

Per risolverlo, alcuni recenti studi, non definitivi, hanno proposto la migrazione continua dei dati al cambiamento dei formati. Ma il cambiamento del formato non è lo stesso della creazione di una nuova copia in quanto richiede l' intervento e la scelta di quali informazioni conservare.

(<http://www.iso/en/ISOOnline.frontpage>)

Ma neanche questo risolve il problema della conservazione in quanto, per J. Ewing, è "virtualmente impossibile salvare ogni pezzo d' informazione dell' ambiente".

A livello organizzativo si sono affermati diversi modelli d' archiviazione dell' informazione scientifica digitale.

I tre modelli organizzativi più diffusi, scrive Richard R. Rowe, sono: l' archiviazione digitale delle case editrici, quella organizzata nei grandi archivi digitali centralizzati (megacentri digitali) e il modello incentrato sulle biblioteche.

¹ Ross, Seamus, (2002), Position paper on integrity and authenticity of digital cultural heritage objects. Round table held in Barcelona on May 6th, 2002, "DigiCULT", v. 3 n. 3, p. 7; Wolpert Ann J. (2002).The future of electronic data. "Nature". Vol. 420, p.17; Rowe Richard R., (2003). Digital archives. Webdebates. "Nature".

² Ewing, John, (2002), Twenty centuries of mathematics:digitizing and disseminating the past mathematical literature. "Notices of the AMS" vol. 49, n. 7, p. 775

³ Ewing, John, (2002), Twenty centuries of mathematics:digitising and disseminating the past mathematical literature. "Notices of the AMS" vol. 49, p. 776.

Il primo modello viene considerato poco praticabile in quanto, scrive Richard R. Rowe, l'archiviazione digitale costituisce un peso finanziario non trascurabile e anche perché non c'è nessuna garanzia dell'esistenza perpetua delle case editrici.⁴

Questo modello viene seguito dalla casa editrice Elsevier che, in collaborazione della biblioteca nazionale olandese, sta creando degli archivi digitali delle sue riviste a stampa e dei preprint.

Un esempio, invece, del secondo modello organizzativo è rappresentato da DSpace (<http://www.dspace.org>) che è un archivio digitale istituzionale sviluppato congiuntamente dal "Massachusetts Institute of Technology Libraries" e dalla Hewlett-Packard Laboratories.⁵

Altri esempi di archivi digitali istituzionali sono gli e-print servers (CogPrints, arXiv.org, PubMed, l'archivio bibliografico della letteratura periodica di biomedicina, BioMed Central, Astrophysical Data System).

Il modello organizzativo incentrato sulle biblioteche e' quello perseguito dalla Biblioteca nazionale olandese che, nel 2003, ha stipulato un accordo con BioMed Central per l'archiviazione digitale di tutti gli articoli pubblicati assicurandone l'accesso perpetuo. Inoltre essa s'impegna ad aggiornare l'archivio digitale man mano che cambiano le tecnologie.⁶

L' "International Congress of Scientific and Technical Information"

(<http://www.icsti.org/icsti.forum/33>) ha discusso molto dei diversi modelli organizzativi per la creazione di un grande archivio digitale centralizzato per la scienza.

Questi dibattiti hanno portato alla elaborazione del progetto denominato "Digital Archives for Science and Engineering Resources" (DESAR) che ha lo scopo di creare un deposito digitale di informazioni scientifiche digitali fondato col modello standard dell' "Open Archives Initiative Protocol" (OAI).

Il progetto ha lo scopo, anche, di permettere lo scambio e la ricerca dei metadati standard fra differenti archivi digitali interoperabili fra loro. Fino al maggio del 2002 c' erano circa 77 archivi digitali che avevano implementato il modello AOI ma il loro numero sta continuamente crescendo. Questo progetto è appoggiato da numerose organizzazioni (università, biblioteche, casa editrici) fra cui la "Digital Library Federation", "Online Computer Library Center" (OCLC).

Un esempio di archiviazione digitale decentralizzata e federativa è costituita da CrossRef che è una associazione "non-profit" di 78 società e case editrici, che attivamente partecipano alla elaborazione di comuni standard di metadati da adattare ai diversi archivi digitali (riviste, libri, letteratura grigia). Il modello aperto scelto da CrossRef è incentrato sul codice identificativo DOI ("Digital Object Identifiers"), che è un identificatore univoco e permanente che può essere incorporato nell' URL.

⁴ Rowe, Richard R., (2003). Digital archives. "Nature" Webdebates, <http://www.nature.com/nature/debates/e-access/Articles/rowe.html>; Cockerill, Matt, (2003), Distributed and centralized technologies. "Nature". Webdebates, <http://www.nature.com/nature/debates/e-access/Articles/cockerill.html>.

⁵ Open Access Now. A journey into Dspace. Biomedcentral. Ultima visita 22 Ottobre 2003 : <http://www.biomedcentral.com/openaccess/archive/>

⁶ Wolpert, Ann J. (2002), The future of electronic data. "Nature" vol. 420, p. 17-18, (07 November 2002); Biomedical. "Online", vol. 27, n. 6, (December 2003), p.12.

La sua strategia organizzativa ed economica, scrive Declan Butler, differisce da quella centralizzata proposte da PubMed Central (PMC), della “US National Library of Medicine”, e da E-Biosci, sviluppato dell’ “European Molecular Biology Organization”(EMBO).⁷

Questi depositi digitali delle istituzioni (università) potrebbero costituire dei modelli per la conservazione a lungo termine delle informazioni digitali.

E’ questo il caso dell’ archivio digitale dei preprint di fisica, matematica arXiv.org (<http://arXiv.org>), creato presso il Los Alamos National Laboratory, e ora conservato dall’ Università di Cornell e da altri siti “mirror” in altre istituzioni mondiali.⁸

Lo studioso, Steven Harnad, è invece fautore della creazione di archivi di “eprint” aperti e interoperabili presso le istituzioni in cui operano gli scienziati capaci di essere inglobati in un archivio globale virtuale.

Alcuni studiosi pensano, invece, che lo sviluppo di standard di comunicazione fra server XML incrementerà lo scambio dei dati fra differenti archivi digitali. L’ uso dei formati dei dati nel formato XML favorirà, scrive M. Cockerill, la conversione dei dati fra i diversi tipi di archivi digitali.⁹

Altri ritengono che l’ introduzione di una nuova generazione di tecnologia web, denominata Web Semantico, incrementerà la comunicazione, estenderà l’ interoperabilità dei database digitali e fornirà 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”).

Il web semantico è stato applicato, scrive lo studioso James Hendler, 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

Gran parte di questi archivi istituzionali digitali usano il CNRI Handle System per memorizzare e renderle accessibile le informazioni digitali. Questo e’ il caso degli archivi digitali DSpace, E-prints che sono conformi ai protocolli dell’ Open Archives Protocol (OAI) e sono interoperabili, accessibili e persistenti.

I matematici

⁷ Declan Butler. (2001). The future of the electronic scientific literature. “Nature”, v.413, p. 1-3 (06 Sept. 2001) ; web debates. <http://www.nature.com/nature/debates/e-access/Articles/opinion2.html>

⁸ Ann J. Wolpert.(2002). The future of electronic data. “Nature”. vol. 420, p. 17-18 (07 November 2002) ; The case for Institutional Repositories: <http://www.arl.org/sparc/IR/ir.html>.

⁹ Cockerill, Matt. (2003). Distributed and centralized technologies:complementary tools to build a permanent digital archive. “Nature”, Webdebates, <http://www.nature.com/nature/debates/e-access/Articles/cockerill.html>, sito visitato 25 nov. 2003; Harnad, Stevan. (2001). Commentary. “Nature”, vo. 410, p. 1024-1025.

¹⁰ Szalay A., Gray J. (2001). The world-wide telescope. “Science”. 21 Sept. 2001; Hendler, James. (2003). Enhanced : Science and the Semantic Web. “Science” vol. 299 (5606), p. 520.

Gli archivi digitali di matematica sono, attualmente, di due tipi: quelli controllati, costituiti dalle riviste elettroniche e libri digitali, e quelli non controllati, formati dagli archivi digitali di preprint (letteratura grigia).

Quest'ultimi sono formati da articoli digitali, rapporti tecnici, relazioni informali non ancora pubblicati nelle riviste o nei libri che ancora non sono stati sottoposti al controllo di qualità tramite il cosiddetto "peer review".

Il più importante è l'archivio di preprint, ArXiv (<http://www.arxiv.org>), creato nel 1991 dal fisico Paul Ginsborg presso il Laboratorio di Los Alamos e ora gestito dall'Università di Cornell, di cui sono state realizzati alcuni "mirror" in diversi luoghi.

Esso è costituito prevalentemente da preprint di fisica e matematica.

Una importante fonte che raccoglie, cataloga e ne permette l'accesso alla letteratura grigia prodotta nei paesi europei è l'archivio SIGLE ("System on Grey Literature in Europe") di cui il 26% (circa) è costituito da preprint di scienze naturali, fra cui quelli di matematica e fisica.¹¹

Fra le fonti digitali controllate, invece, ci sono i libri e le riviste i cui principali archivi digitali (alcuni sono per ora progetti, alcuni sono a pagamento e altri sono ad accesso libero) sono:

JSTOR (Journal STORAGE) (<http://www.jstor.org>) (archivio digitale interdisciplinare in cui ci sono, per ora, 17 riviste digitalizzate di matematica pubblicate dall' American Mathematical Society e di altre società; l' archivio si sta estendendo);

DIEPER (Digitised European Periodicals) (<http://dieper.aib.uni-linz.ac.at>) digitalizzazione della rivista "Monatshefte für Mathematik und Physik" (1890-1918); Biblioteka Wirtualna Matematyki (<http://matwbn.icm.edu.pl>), comprende le riviste Fundamenta Mathematicae, 1920-1993, Studia Mathematica (1929-1964), Prace matematyczno-fizyczne (1888-1952) e alcuni libri elettronici di matematici polacchi;

Departamento de Ingeniería Matemática Universidad de Chile

(<http://www.dim.uchile.cl/revmat.html>) (Revista de Matemáticas Aplicadas, 1994-2002;

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" (v. 1-68, 1868-1942) con la creazione di un archivio digitale e con legami a libri digitali del progetto "Digital Library" della Cornell University e a quelli del progetto "Gallica"; il progetto "Gallica" (Bibliothèque nationale de France, <http://gallica.bnf.fr>) comprende la digitalizzazione delle seguenti riviste: Journal de Mathématiques Pure et Appliquées, 1936-1880, Comptes Rendus Hebdomadaires des Séances de l' Académie des Sciences, 1835-1930;

NUMDAM (Numérisation de Documents Anciens Mathématiques, <http://www.numdam.org>) comprende le seguenti riviste: 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 dérivées partielles, 1974-2000, Mémoires de la Société Mathématique de France, 1964-1992, Publications Mathématiques de l'IHES, 1959-1997; progetto Euclid (<http://ProjectEuclid.org/mmi>), comprende la rivista Michigan Mathematical Journal, dal 1952 a oggi; SciELO (Scientific Electronic Library Online, <http://www.scielo.cl/scielo.php>), comprende la rivista: Proyecciones – Revista de Matemática, 2000-2002;

WDML-Göttingen (<http://Gottinger> 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 di altre case editrici.¹²

¹¹ Bardelloni, Angelo (1996). Preprint elettronici nel campo della matematica pura ed applicata. In : La letteratura grigia 2° Convegno Nazionale. Roma : Istituto Superiore di Sanità, p. 202-206; Wood, David N., Smith, Andrei W. (1996). EAGLE: a model for international cooperation. In : La Letteratura grigia 2o Convegno Nazionale. Roma : Istituto Superiore di Sanità, p. 93-96.

Questi diversi archivi non sono coordinati fra loro e l'obiettivo del progetto DML e' quello di creare un coordinamento fra loro e i due database bibliografici di matematica, MathSciNet e Zentralblatt MATH.

Questo legame e' stato creato fra MathSciNet, e i documenti elettronici conservati nell' archivio JSTOR.

Una operazione simile è stata fatta anche fra le citazioni bibliografiche dell' archivio bibliografico Zentralblatt MATH e gli articoli conservati nell' archivio del progetto WDML-Gottingen.

Ultimamente sono stati ridigitalizzati (dal 1868 al 1942) tutti gli articoli del "Jahrbuch uber die Fortschritte der Mathematik" come parte del progetto tedesco "Electronic Research Archiv for Mathematics"(ERAM) (<http://www.emis.de/ZMATH/>).

In questo modo il database, Zentralblatt MATH, rende accessibile la letteratura di matematica pubblicata dal 1868 a oggi.

Per i libri digitali di matematica è stato realizzato un catalogo centralizzato disponibile presso l' Università Joseph Fourier di Grenoble ("The Cellule MathDoc") che censisce i libri di quattro siti web: il "Digital Math Books", della biblioteca dell' Università di Cornell, i libri di matematica della collezione "Gallica" della Biblioteca Nazionale di Francia, i libri digitali del progetto "WDML-Gottingen", presso il "Gottinger Digitalisierungszentrum", e quelli della collezione storica dell' Università del Michigan (<http://math-doc.ujf-grenoble.fr/LiNuM/>).

Recentemente a un convegno, tenuto presso l' università di Gottingen, un ristretto gruppo del "Mathematical Reviews", del "Zentralblatt fur 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 (MathSciNet e Zentralblatt MATH) e gli archivi digitali delle riviste di matematica.

Recentemente, col progetto LIMES, finanziato dalla Commissione Europea e dall' istituto di fisica FIZ Karlsruhe, si cerca di estendere, a livello europeo, questo database bibliografico digitale di matematica.

I fisici

A differenza dei matematici, i fisici sono stati i precursori nella creazione dei primi archivi elettronici e delle prime riviste elettroniche. Perciò sono stati i fisici i più sensibili interpreti dell'esigenza di creare archivi digitali aperti con standard condivisi con cui conservare la loro letteratura.

E' 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.

Questa sensibilita' non era presente nel 1991 quando il fisico, 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, astrofisica, matematica, che divenne, ben presto, uno dei principali strumenti di circolazione dell' attività di ricerca di letteratura grigia fra i fisici.

Attualmente questo archivio di letteratura grigia è ospitato presso la biblioteca della Cornell University.¹³

¹² Allyn Jackson. (2003), The Digital Mathematics Library. "Notices of the AMS", vol. 50, n. 8, p.920

¹³ Quirin Schiermeier. (2000). "Nature", vol. 408, p. 757 (14 December 2000); Stevan Harnad. (1999). Correspondence. "Nature", vol. 401, p. 423 (30 September 1999).

Ma gli articoli nell' archivio ArXiv (<http://www.arxiv.org>) non sono controllati ("peer-reviewed") e questo fatto costituisce un limite nella loro accettazione come fonte principale nelle citazioni e nella conservazione della letteratura di fisica.

Ma questo modello di comunicazione scientifica ha oscurato i confini fra la letteratura grigia (rappresentata dai preprint) e quella controllata pubblicata nelle riviste.

Di questo archivio digitale di preprint si sono creati, per motivi di sicurezza e di accesso, delle copie digitali ("mirror site") uguali in diversi posti, fra cui una in Italia, presso la SISSA di Trieste (<http://www.babbage.it>). Esse vengono continuamente aggiornate.

Anche il CNRS ("Centre National de la Recherche Scientifique") francese, nel 2000, ha deciso di creare una sezione francese dell' archivio digitale di preprint di fisica di Los Alamos.¹⁴

Il formato dei documenti di questi archivi digitali sono memorizzati in formato TeX e in altri formati (PostScript, LateX, PDF)

Nel 2002 l' archivio digitale dei preprint 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 da più di 77 registrati archivi digitali (maggio 2002) ma il loro numero sta continuamente crescendo.

Il protocollo OAI è stato anche adottato dal CERN Document server software per la ricerca di 550000 record di metadati di documenti digitali prodotti dai fisici del CERN di Ginevra.¹⁵

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' "Open Archives Initiative" come strumento base di comunicazione fra il portale e i sottostanti servizi (archivi, servizi, riviste elettroniche) operando in modo trasparente.¹⁶

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 permette la ricerca bibliografica di più di 500.000 articoli di fisica, compresi articoli di riviste, preprints, rapporti tecnici, atti di congressi e tesi.

Questi sono indicizzati dalla biblioteca universitaria di Stanford e da DESY a partire dal 1974.

Di questo archivio sono stati costruiti dei siti "mirror" in Gran Bretagna, Giappone, Russia e Germania.¹⁷

¹⁴ Susan Y. Crawford (et al.). (1996). From print to electronic. Medford, NJ : Information Today, p. 69-75; Butler Declan. (2000). French take physics archives into the future. "Nature".v. 407, p. 825.

¹⁵ Michael Nelson (2002). New developments in OAI (OAI-PMH2). Workshop Report 1, Open Archives Forum (14 August 2002); Martin Vesely. CERN Document Server Software.

¹⁶ Bertocco, Sara. (2001).Torii, an Open Portal over Open Archiv, "High Energy Physics Libraries Webzine", issue 4, June 2001 (URL:<<http://library.cern.ch/HEPLW/4/papers/4/>>)

¹⁷ Susan Y. Crawford (et al.). (1996). From print to electronic. Medford, NJ : Information Today, p. 69-70.

Ci si è resi conto della disomogeneità dei diversi archivi digitali di fisica e per questo diversi studiosi hanno proposto l'interoperabilità fra le vari documenti digitali, fra gli e-prints, fra gli archivi digitali distribuiti e le riviste elettroniche.

Lo strumento standard che lo permette è il protocollo standard OAI-PMH adottato da diversi archivi digitali fra cui quello di astrofisica, ADS, e quello di preprint di fisica di Los Alamos e della Cornell University.

Gli astronomi hanno creato una rete di archivi digitali distribuiti in rete, incentrati sul database bibliografico "The NASA Astrophysics Data System" (ADS - <http://adswww.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.¹⁸

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 esterne come articoli di riviste a testo completo, cataloghi e archivi.

Un altro esempio è il "bibcode" (la descrizione delle referenze pubblicate con 19 caratteri) che è stato adattato dal principale database bibliografico dell'astronomia, ADS, e 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 in rete e di renderlo interoperabile con lo standard DOI.

Una simile funzionalità dell'ADS si ha con SLAC-SPIRES (<http://www.slac.stanford.edu/spires>) gli archivi digitali di fisica delle alte energie.

Un altro importante archivio bibliografico di fisica è "Science Abstract" prodotto dall'"Institute of Electrical Engineers"(IEE).

Dalla primavera del 2004 questo archivio digitale bibliografico sarà disponibile anche in formato digitale. Inspec ha digitalizzato tutta la collezione cartacea (1898-1968) per poterla meglio consultare e conservare. È stato creato un archivio digitale in formato XML comprendente 800.000 record (circa).

I principali archivi digitali di fisica sono, però, quelli delle riviste elettroniche.

Uno di questi è rappresentato dall'archivio digitale PROLA (Physical Review Online Archive) dell'"American Physical Society"(APS).

Gli altri sono quelli delle principali case editrici scientifiche (Springer, IOP, European Physical Society, Elsevier, World Scientific, American Institute of Physics e altre).

Non è ancora molto chiaro se questi archivi digitali verranno gestiti dalle case editrici oppure dalle biblioteche o a consorzi.

La Elsevier, ad esempio, ha fatto un accordo con la Biblioteca Nazionale d'Olanda per la conservazione a lungo termine dei suoi archivi digitali di riviste.

Alcune case editrici private hanno, invece, assegnato la gestione dei loro archivi digitali di riviste a società indipendenti, come la Highwire Press, mentre altre li gestiscono in proprio.

Il PROLA è il principale strumento bibliografico per l'accesso alle riviste elettroniche pubblicate dall' American Physical Society (APS).

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 depositato presso la Library of Congress e un' altra ("mirror") presso la biblioteca dell' Università di Cornell (<http://cornell.mirror.aps.org/info>).

¹⁸ Genova Françoise. (2002). Online information in astronomy:from networking to a virtual observatory. In Open Archives Forum, (14 August 2002)

La maggior parte della collezione del PROLA è costituita da immagini scannarizzate, in formato immagine GIF o in formato PDF, delle riviste a stampa. Le informazioni bibliografico 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 documento originali.

Nell' ottobre del 2003 il "Los Alamos National Laboratory" ha ottenuto la licenza, dall' "American Physical Society", di conservare l' intera collezione digitale delle sue riviste e l' archivio PROLA. APS e Los Alamos useranno i protocolli dell' "Open Archives Initiative" per mantenere la copia sincronizzata con la versione originale dell' archivio.

I due enti si impegnano, inoltre, di usare un modello standard di distribuzione del contenuto degli archivi usando l' "Open Archives Protocol for Metadata Harvesting".

Il modello del preprint si fisica di Los Alamos è stato poi imitato anche dalle scienze cognitive e ultimamente anche dalla biologia (<http://arXiv.org/archive/q-bio>).

In quest' ultimo caso sono state sollevate, da diversi studiosi, delle fondate preoccupazioni sulla accessibilità delle informazioni mediche non controllate da parte di medici e pazienti.¹⁹

Accanto a questo modello di autoarchiviazione libera e non controllata dei risultati della ricerca, i fisici hanno creato le loro prime riviste completamente elettroniche.

Queste costituiscono la fonte principale per l' archiviazione digitale, a lungo termine, della letteratura controllata in fisica.

Sono stati i fisici della SISSA di Trieste a creare nel 1998 la loro rivista completamente elettronica chiamata JHEP ("Journal of High Energy Physics" (<http://www.iop.org/journals/jhep> o <http://jhep.sissa.it>).

Tutta l'attività editoriale, dalla sottomissione alla pubblicazione, e' stata digitalizzata attraverso un software robot appositamente costruito. L'iniziativa iniziale e' stata finanziata con fondi CEE.

Di questa rivista elettronica si sono creati, per motivi di sicurezza e d' integrità delle informazioni digitali, due archivi digitali: uno presso la casa editrice IOP e l' altro, di riserva, presso la SISSA.

Da questo modello è nata nel 2002 la rivista elettronica di astrofisica "Journal of cosmology and astroparticle physics" (JCAP, <http://www.iop.org/journals/jcap>). Un' altra nuova rivista elettronica è in fase di creazione.

Il modello di JHEP ha sollevato, quasi subito, il problema della conservazione delle informazioni digitali, della accessibilità e dell' integrità dei documenti scientifici digitali.

Quest' ultimo problema è stato risolto dotando gli articoli digitali, più recenti, di metadati standard (nel formato del Dublin Core set) e conservandoli in due posti diversi: uno presso la casa editrice IOP e una copia presso la SISSA.

La questione, invece, della conservazione a lungo termine di queste riviste digitali si è pensato di risolverlo con la loro memorizzazione dei documenti digitali in formati standard come quelli proposti dall' "Open Archives Initiative" (<http://www.openarchives.org>) e nel formato PDF.

In campo accademico un altro importante archivio digitale, non solo di fisica ma di scienza, è quello realizzato dalla casa editrice Highwire Press.

Le prospettive future

Uno dei più importanti progetti di digitalizzazione di matematica e' il "The Digital Mathematics Library" (DML), che ha l' obiettivo di creare una fonte digitale centrale di tutta la letteratura

¹⁹ Declan Butler. (2003). Biologists join physics preprint club. "Nature" vol. 425, p. 548, (09 October 2003).

matematica mondiale prodotta accessibile a chiunque in possesso di computer e a una connessione Internet.

L' ideatore di questo ambizioso progetto è il matematico americano John H. Ewing, direttore esecutivo dell' "American Mathematical Society" (AMS), e l'AMS.

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 questa ampia impresa.²⁰

Questo progetto americano e' stato imitato da un gruppo di matematici europei che, nell'aprile del 2003, ha proposto alla Commissione Europea un progetto di finanziamento quinquennale per la realizzazione di un DML in Europa (denominato DML-EU) capace di realizzare la digitalizzazione delle riviste di matematica europee a livello nazionale.

Fra le principali difficoltà riscontrate nella sua realizzazione, scrive John Ewing, 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 ai formati di memorizzazione e di archiviazione di lungo termine.

Questi nodi possono essere superati, scrive il matematico John Ewing, con l' odierna tecnologia e anche perché la matematica è una disciplina ideale per la retro-digitalizzazione.

Diversamente dalle altre discipline scientifiche, scrive il matematico Allyn Jackson, la matematica è "inestinguibile dalla sua letteratura" e i matematici usano molto la letteratura passata mentre lavorano alla frontiera della ricerca.²¹

Seguendo i suggerimenti di queste due studiosi, diverse società accademiche, biblioteche, case editrici hanno iniziato a digitalizzare la letteratura passata di matematica.

Nel campo della fisica, invece, si auspica la realizzazione di legami fra gli archivi digitali e quelli bibliografici in modo da permettere il passaggio digitale dalla ricerca bibliografica al testo completo dei documenti digitali.

Anche qui' i problemi da superare sono quelli elencati precedentemente dal matematico John Ewing.

²⁰ John H. Ewing. (2002). Twenty centuries of mathematics:digitizing and disseminating the past mathematical literature. "Notices of the AMS". vol.49, n. 7, p. 771-7.

²¹ Allyn Jackson. (2003), The Digital Mathematics Library. "Notices of the AMS", vol. 50, n.8, p. 919.