

Aspetti della preservazione digitale

Fabrizia Bevilacqua

Dipartimento di Filologia classica e medievale
Università degli studi di Parma
fabrizia.bevilacqua@unipr.it

Il caso degli archivi istituzionali

Presso diverse università e istituti di ricerca in Italia sono attualmente in corso di sviluppo progetti per la realizzazione di archivi istituzionali di materiali digitali.¹ Con il termine archivio istituzionale (*institutional repository*) si intende un insieme di servizi per la gestione e disseminazione dei materiali digitali prodotti dalla comunità accademica o di ricerca. Questi servizi comprendono non solo l'organizzazione dell'accesso e la distribuzione attraverso Internet, ma anche, nei casi opportuni, la preservazione a lungo termine.² Inizialmente concepiti come strumenti per favorire la libera diffusione degli articoli di ricerca, gli archivi istituzionali tendono ora a considerare una più ampia varietà di tipologie di oggetti digitali, come tesi, file video e sonori, immagini, materiali didattici, dati amministrativi e collezioni di dati numerici. L'attuale sviluppo di questi strumenti può essere considerato per molti aspetti ancora in fase iniziale, dal momento che gli archivi ospitano limitate quantità di materiali e che occorrerà molto tempo per promuovere efficacemente il servizio presso la comunità accademica. Tuttavia, non deve sembrare prematuro iniziare a pensare fin da ora alle problematiche relative alla preservazione a lungo termine dei contenuti. È anzi opportuno che le istituzioni si preparino a definire, a cominciare dalle fasi progettuali, se assumere l'impegno di preservare i materiali prodotti dalla co-

munità e depositati nell'archivio digitale, nonché naturalmente di garantire nel tempo l'esistenza del servizio stesso. Per farlo è essenziale comprendere gli aspetti chiave della preservazione digitale, valutare le risorse e le competenze richieste e scegliere le strategie più efficaci per gestire un progetto di preservazione. In primo luogo: la preservazione rappresenta un obiettivo da perseguire, oppure è più opportuno concentrarsi su altri aspetti prioritari, come ad esempio la promozione, l'organizzazione del materiale e la valutazione di qualità? Quali sono le implicazioni, i costi e i benefici di un impegno per la preservazione? Si tratta di un progetto sostenibile anche per le università di piccole dimensioni?

Finalità e strumenti della preservazione

I possibili motivi per decidere di preservare il proprio archivio istituzionale sono diversi. Per esempio per garantire l'accesso aperto alla produzione digitale; perché i materiali depositati sono conosciuti e citati da altri membri della comunità accademica; perché vi sono documenti di notevole valore per l'istituzione o disponibili solo attraverso l'archivio; perché la prospettiva di una preservazione a lungo termine può costituire un elemento di attrattiva per gli autori.³ Secondo Paul Wheatley, autore di un rapporto sulla preservazione

digitale nel contesto degli archivi istituzionali redatto per la Digital Preservation Coalition (DPC), gli obiettivi funzionali della preservazione digitale sono:

- 1) garantire l'integrità e autenticità dei dati;
- 2) garantire che i dati siano reperibili ed estraibili da parte degli utenti;
- 3) garantire che i dati siano comprensibili e interpretabili;
- 4) assicurare i primi tre obiettivi a lungo termine.⁴

Alcuni dei software open source attualmente disponibili per la gestione degli archivi istituzionali offrono strumenti a supporto della preservazione e in particolare dei primi due obiettivi citati, i quali si riferiscono ad una preservazione di base e non contengono immediate implicazioni a lungo termine.⁵ Tra questi, DSpace soprattutto pone gli aspetti della preservazione digitale al centro delle prospettive di sviluppo del sistema e offre specifiche funzioni a supporto della gestione di questi aspetti.⁶

In particolare:

- il modello di licenza di distribuzione sottoposto agli autori che sottomettono contributi all'archivio chiede espressamente il consenso all'effettuazione di una migrazione ad altro formato o supporto, nonché alla realizzazione di più copie con lo scopo di assicurare la sicurezza e la preservazione;
- sono definiti tre tipi di supporto rispetto ai formati, distinguendo tra formati per cui si garantisce un supporto completo da parte del

servizio, formati riconosciuti ma per cui non si garantisce un supporto completo, e infine formati non riconosciuti e per cui non si garantisce alcun tipo di supporto;

- il sistema effettua un controllo sull'integrità dell'oggetto digitale basato sulla somma di controllo (*checksum*);

- DSpace attribuisce all'oggetto archiviato un identificativo permanente, al fine di garantire che esso sia sempre identificabile e reperibile anche quando ne viene cambiata la localizzazione fisica.

Tuttavia, al momento DSpace non offre strumenti di gestione delle funzioni di preservazione a lungo termine (la cosiddetta preservazione funzionale), implicite nel terzo e naturalmente nel quarto degli obiettivi citati da Wheatley, in grado di supportare le istituzioni nella gestione di un piano di preservazione attraverso cui superare l'obsolescenza tecnologica e dei supporti fisici, assicurando che il materiale archiviato continui a essere decodificabile e leggibile nel tempo.⁷ Le principali limitazioni esistenti al momento consistono nell'insufficiente disponibilità di metadati per la preservazione, nonché nella mancanza di strumenti per identificare i rischi di obsolescenza di specifici formati di file e per gestire una strategia di preservazione di tali formati.⁸

Gli sviluppatori di DSpace insieme a esperti dell'Università di Cambridge e ad altri utilizzatori del software stanno lavorando per migliorare questi aspetti e si propongono in particolare di:

- sviluppare strumenti per la gestione delle funzioni di migrazione immediata e automatizzata (*migration on ingest*) e migrazione a richiesta;
- definire modelli per la registrazione di metadati specifici per la preservazione;

- predisporre dei protocolli d'azione pronti all'uso per la preservazione di specifici formati;

- realizzare un'analisi della distribuzione dei costi nell'ambito delle strategie di preservazione.⁹

Aspetti politici e organizzativi

Se diverse problematiche tecniche della preservazione digitale cominciano ad essere affrontate nei programmi di sviluppo dei software di gestione, esistono anche aspetti organizzativi, politici e gestionali relativi alla preservazione che le università attive o ancora solo orienta-

tori di archivi istituzionali si sono incontrati per discutere le problematiche della preservazione nel contesto degli archivi istituzionali, confrontare le strategie in atto e definire possibili iniziative a supporto delle istituzioni impegnate nella gestione di tali servizi.¹⁰ L'orientamento generalmente condiviso è che il problema della preservazione non debba preoccupare troppo chi si trova in una fase ancora iniziale del servizio, in cui sono prioritari gli aspetti organizzativi e di promozione nei confronti di professori e ricercatori. Una volta che l'archivio è popolato da una consistente quantità di materiale, diventa invece più importante definire le linee d'azione e pianificare le risorse necessarie per preservare i materiali depositati.¹¹ Tuttavia, la preservazione costituisce un aspetto da non trascurare nelle fasi iniziali del progetto, sia perché esso riguarda strettamente gli obiettivi strategici che ci si propone di raggiungere attraverso l'implementazione di un archivio istituzionale, sia perché attivare alcune strategie dalle fasi iniziali della vita di uno strumento elettronico può rendere più semplice ed economica la sua successiva manutenzione. Si tratta quindi di preparare la preservazione futura dell'archi-

vio più che di gestire la preservazione del materiale esistente.¹² Ecco alcuni aspetti che le università dovranno prendere in considerazione al fine di definire le politiche di preservazione a lungo termine dei propri archivi istituzionali:

- pianificare la longevità dell'archivio stesso, cercando di assicurarsi fonti stabili di finanziamento che garantiscano la continuità del servizio;
- definire le responsabilità dell'istituzione in rapporto alla preservazione dei materiali;



te verso l'implementazione di un archivio istituzionale stanno iniziando a discutere e ad affrontare. Ogni decisione riguardo all'opportunità di sviluppare un piano di preservazione deve infatti essere guidata da definiti obiettivi strategici da parte dell'istituzione, non dalla semplice disponibilità di specifiche tecnologie. In occasione di un convegno organizzato nello scorso ottobre a Londra dalla DPC e dal Consortium of Research Libraries in the British Isles (CURL), diversi ge-

- scegliere il software più rispondente alle proprie esigenze di preservazione;
- valutare le necessità di preservazione in relazione alle tipologie dei materiali depositati e dei formati;
- valutare i costi e i benefici rispetto alle dimensioni dell'organizzazione e ai suoi obiettivi strategici;
- valutare i tempi della preservazione in relazione alle tipologie di materiale, distinguendo per esempio tra preservazione a medio termine e a lungo termine per diverse tipologie.

Il problema dei costi della preservazione nel contesto degli archivi istituzionali è ancora di difficile definizione, anche perché non esistono ancora chiari modelli di costo per gli archivi istituzionali in generale: tuttavia, cercare di ottenere fonti di finanziamento affidabili e stabili può costituire un primo passo per essere in grado di affrontare con il tempo anche un piano di preservazione.¹³ Le scelte riguardo alle tipologie di materiali e ai formati da preservare devono essere gestite con attenzione, facendo riferimento agli obiettivi strategici che si intendono perseguire nel tempo, alla natura dell'archivio, alle risorse disponibili, nonché alle possibili ripercussioni sugli utenti. Infatti, una politica eccessivamente rigida riguardo ai formati accettati dal deposito, anche se presa per garantire la preservazione a lungo termine, potrebbe avere la conseguenza di scoraggiare i ricercatori dal sottomettere materiali, se non affiancata da un efficiente servizio di supporto alla conversione dei file. Le scelte delle varie istituzioni riguardo a tale aspetto possono essere anche molto differenti l'una dall'altra. Così per esempio all'Università di Edinburgh, che gestisce un archivio di e-print e tesi, gli utenti sono incoraggiati a scegliere formati per i quali si garantisce un supporto completo rispetto alla preservazione, mentre all'Universi-

tà di Cambridge, il cui deposito ospita differenti tipologie di oggetti digitali e conta una limitata quantità di articoli di ricerca e una molto maggiore presenza di file video e immagini, si è deciso di incoraggiare l'invio di materiale in una varietà di formati: questa scelta ha comportato un conseguente maggiore impegno per lo sviluppo di un'infrastruttura per la preservazione a lungo termine.¹⁴ Gli utenti stessi hanno esigenze di preservazione di cui occorre tenere conto. Un'indagine effettuata all'Università di Parma tra professori e ricercatori della Facoltà di lettere e filosofia, al fine di identificare le esigenze degli utenti rispetto all'implementazione di un archivio istituzionale, ha rivelato per esempio che l'utenza accademica ha forti aspettative riguardo alla preservazione dell'integrità e autenticità dei documenti elettronici attraverso un deposito istituzionale.¹⁵ Il tipo di supporto offerto da DSpace, il software selezionato all'Università di Parma, sembra poter soddisfare queste esigenze di sicurezza. Qualunque siano le politiche assunte riguardo alla preservazione dell'archivio, è importante che gli utenti siano informati delle scelte e delle implicazioni connesse: l'esistenza di strumenti di garanzia nei confronti dell'integrità e persistenza dei documenti può costituire un elemento decisivo per convincere i ricercatori a passare da sistemi più aleatori di distribuzione dei propri materiali attraverso il web al deposito nell'archivio digitale della propria organizzazione.¹⁶

Conclusioni

L'adozione di strategie per una preservazione di base dei materiali depositati, anche se non costituisce necessariamente un aspetto indispensabile per tutti gli archivi istituzionali, sembra essere un obiettivo sostenibile per molte isti-

tuzioni anche di piccole dimensioni. Queste strategie potranno per esempio includere la selezione di tipologie di materiali da preservare, l'individuazione di formati privilegiati, l'eventuale attivazione di un servizio di supporto agli autori per la conversione dei file, l'implementazione dell'identificativo permanente e l'effettuazione di copie di backup con archiviazione esterna. È prevedibile, tuttavia, che non tutte le università saranno in grado di affrontare da sole le complesse problematiche legate alla preservazione a lungo termine degli archivi istituzionali, né a sostenerne singolarmente i costi che potrebbero anche essere molto elevati. Così come per altri settori della preservazione digitale, sarà essenziale sia imparare dall'esperienza delle prime istituzioni che stanno sperimentando infrastrutture per la preservazione e adottarne le buone pratiche, sia attivare iniziative cooperative con la partecipazione di diverse università. In questa direzione si sta muovendo il progetto SHERPA che, nato con lo scopo di studiare e promuovere la creazione di archivi istituzionali nelle università inglesi,¹⁷ ha lanciato un'iniziativa proprio sui problemi della preservazione nel contesto degli archivi istituzionali. Il progetto, chiamato SHERPA DP Project, avrà i seguenti obiettivi: sviluppare un'infrastruttura permanente per la preservazione basata sul modello OAIS;¹⁸ realizzare una *Digital repository preservation user guide*, in cui saranno suggeriti standard, buone pratiche, protocolli e processi da adottare nella gestione di archivi di e-print; esplorare l'utilizzo di tecnologie open source per migliorare ed estendere le funzionalità dei software di gestione degli archivi istituzionali rispetto alla preservazione; infine, esplorare le possibilità di un approccio distribuito alla preservazione, che preveda la distinzione di un livello di

accesso ai contenuti, gestito dai vari archivi istituzionali, e di un livello di archiviazione per la preservazione a lungo termine, gestito da un'agenzia sovra-istituzionale.¹⁹ Questo approccio è stato scelto per ovviare alla mancanza di strumenti avanzati per la preservazione in software come DSpace e Eprints, e alla scarsità di personale universitario con competenze specifiche riguardo a tali aspetti; inoltre si cerca di sviluppare una strategia che, eliminando la necessità di gestire servizi ripetitivi nelle singole istituzioni, possa favorire un risparmio nei costi.

Sarà importante seguire con attenzione l'evoluzione di questo progetto, i cui lavori dovrebbero concludersi alla fine del 2006, nonché le iniziative intraprese dal gruppo di lavoro della Federazione degli utilizzatori di DSpace per migliorare le funzionalità a supporto della preservazione a lungo termine.²⁰ Tuttavia non è forse prematuro iniziare a confrontarsi anche in Italia su queste problematiche, provando a mettere in comune le esperienze e i punti di vista, nonché ad avanzare ipotesi di lavoro al fine di sviluppare un approccio coordinato al problema della preservazione nel contesto degli archivi istituzionali, esplorando possibili ruoli, strategie e livelli di responsabilità con il coinvolgimento non solo delle università ma anche delle altre istituzioni interessate agli aspetti della preservazione digitale.

Note

¹ Secondo Paola Gargiulo sono attualmente attivi già dieci archivi istituzionali, mentre tre sono in fase di sperimentazione. Cfr. P. GARGIULO, *Introduzione alle problematiche dell'Open Access in Italia*, intervento presentato al convegno "Gli atenei italiani per l'Open Access: verso l'accesso aperto alla letteratura di ricerca", Messina, 4-5 novembre 2004, <[\[pic.it/conf/viewabstract.php?id=35&cf=1\]\(http://www.ae-pic.it/conf/viewabstract.php?id=35&cf=1\)> \(novembre 2004\). Sullo sviluppo degli archivi di e-print in Italia si veda anche: V. COMBA, *Il progetto AEPIC \(Academic E-Publishing Infrastructures - CILEA\): le esigenze degli autori e lo stato dell'arte dei servizi*, "Bibliotime", 6 \(2003\), 3, <<http://www.spbo.unibo.it/bibliotime/num-vi-3/comba.htm>>; A. DE ROBBIO, *Gli archivi e-prints in Italia*, "Bibliotime", 7 \(2004\), 1, <<http://www.spbo.unibo.it/bibliotime/num-vii-1/derobbio.htm>>; G. VITIELLO, *La comunicazione scientifica e il suo mercato*, "Biblioteche oggi", 21 \(2003\), 5, p. 49-57, <<http://www.bibliotecheoggi.it/2003/20030503701.pdf>>.](http://www.ae-</p></div><div data-bbox=)

² R. CROW, *The case for institutional repositories: a SPARC position paper*, SPARC, 2002, <<http://www.arl.org/sparc/IR/ir.html>>; C.A. LYNCH, *Institutional repositories: essential infrastructure for scholarship in the digital age*, "ARL Bimonthly Report", (2003), 226, <<http://www.arl.org/newsltr/226/ir.html>>; A. SMITH, *New-model scholarship: how will it survive?*, CLIR, 2003, <<http://www.clir.org/pubs/abstract/pub114abst.html>>; A.M. TAMMARO - T. DE GREGORI, *Ruolo e funzionalità dei depositi istituzionali*, "Biblioteche oggi", 22 (2004), 10, p. 7-19.

³ Per un esame delle motivazioni contrarie e a favore della preservazione degli archivi di e-print si veda: H. JAMES, et al., *Feasibility and requirements study on preservation of e-prints: report commissioned by the Joint Information Systems Committee (JISC)*, JISC, 2003, p. 20-21, <http://www.jisc.ac.uk/uploaded_documents/e-prints_report_final.pdf>; S. PINEFIELD - H. JAMES, *The digital preservation of e-prints*, "D-Lib Magazine", 9 (2003), 9, <<http://www.dlib.org/dlib/september03/pinfield/09pinfield.html>>.

⁴ P. WHEATLEY, *Institutional repositories in the context of digital preservation*, DPC Technology Watch Series Report 04-02, 2004, <<http://www.dpconline.org/docs/DPCTWf4word.pdf>>, (novembre 2004).

⁵ Una guida aggiornata ai software open source per la gestione degli archivi istituzionali è disponibile sul sito della Budapest Open Access Initiative: OPEN SOCIETY INSTITUTE, *A guide to institutional repository software*,

<<http://www.soros.org/openaccess/software/>>, (novembre 2004).

⁶ DSPACE FEDERATION, *DSpace features*, <<http://dspace.org/technology/features.html>>, (novembre 2004).

⁷ N. BEAGRIE - M. JONES, *Preservation management of digital materials: a handbook*, London, The British Library, 2001; disponibile anche online: <<http://www.dpconline.org/graphics/handbook/>>; D. BEARMAN, *Reality and chimeras in the preservation of electronic records*, "D-Lib Magazine", 5 (1999), 4, <<http://www.dlib.org/dlib/april99/bearman/04bearman.html>>; P. CONWAY, *Preservation in the digital world*, CLIR, 1996, <<http://www.clir.org/pubs/reports/conway2/index.html>>; CPA/RLG TASK FORCE ON ARCHIVING OF DIGITAL INFORMATION, *Preserving digital information: report of the Task Force on Archiving of Digital Information*, RLG, 1996, <<http://www.rlg.org/ArchTF/tfadi/index.htm>>, (novembre 2004); J. JACKSON, *Digital longevity: the lifespan of digital files*, DPC, <<http://www.dpconline.org/graphics/events/digital-longevity.html>>, (novembre 2004); D. ROTHENBERG, *La conservazione dei documenti digitali*, "Le scienze", 27 (1995), 319, p. 16-21.

⁸ J. JACKSON, *Digital longevity: the lifespan of digital files*, cit.; G.W. LAWRENCE, *Risk management of digital information: a file format investigation*, CLIR, 2000, <<http://www.clir.org/pubs/reports/pub93/contents.html>>; A. SALARELLI - A.M. TAMMARO, *La biblioteca digitale*, Milano, Editrice Bibliografica, 2000, p. 167-168; A. STANESCU, *Assessing the durability of formats in a digital preservation environment: the INFORM methodology*, "D-Lib Magazine", 10 (2004), 11, <<http://www.dlib.org/dlib/november04/stanescu/11stanescu.html>>; P. WHEATLEY, *Survey and assessment of sources of information on file formats and software documentation: final report*, JISC, 2003, <http://www.jisc.ac.uk/uploaded_documents/FileFormatsreport.pdf>.

⁹ Id., *A way forward for developments in the digital preservation functions of DSpace: options, issues and recommendations*, 2003, <<http://www.dspace.org/news/articles/DpAndDSpace.pdf>>, (novembre 2004); Id., *Preserv-*

ing our past: a joint digital repository project for University and MIT, University of Cambridge, 10 febbraio 2003, <<http://www.admin.cam.ac.uk/news/press/dpp/2003021001>>, (novembre 2004).

¹⁰ CURL/DPC Forum "Digital preservation in institutional repositories", London 19th October 2004, <<http://www.dpconline.org/graphics/events/041019forum.html>>, (novembre 2004).

¹¹ Per esempio all'Università di Edinburg il problema della preservazione è stato affrontato quando l'archivio comprendeva alcune centinaia di oggetti digitali: J. McCOLL, *DSpace institutional repositories and digital preservation*, intervento presentato al CURL/DPC Forum "Digital preservation in institutional repositories", cit.

¹² H. JAMES, et al., *Feasibility and requirements study on preservation of e-prints*, cit.; B. LAVOIE – L. DEMPSEY, *Thirteen ways of looking at... digital preservation*, "D-Lib Magazine", 10 (2004), 7/8, <<http://www.dlib.org/dlib/july04/lavoie/07lavoie.html>>; S. PINEFIELD – H. JAMES, *The digital preservation of e-prints*, cit.

¹³ *Ibidem*.

¹⁴ J. McCOLL, *DSpace institutional repositories and digital preservation*, cit.; J. DOWNING, *Planned digital preservation R&D*, intervento presentato al CURL/DPC Forum "Digital preservation in institutional repositories", cit.

¹⁵ P. GOZZETTI, *Indagine sulle aspettative dell'utenza scientifica per i servizi di DSpace: il caso della Facoltà di lettere dell'Università di Parma. Tesi di laurea, a.a. 2003-2004*, p. 138, <<http://hdl.handle.net/1889/97>>. Riguardo al problema dell'autenticità e integrità delle risorse digitali nell'ambito della ricerca scientifica si veda: D. BARMAN – J. TRANT, *Authenticity of digital resources: towards a statement of requirements in the research process*, "D-Lib Magazine", June 1998, <<http://www.dlib.org/dlib/june98/06barman.html>>.

¹⁶ M. SMITH, *DSpace for e-print archives*, "High Energy Physics Libraries Webzine", 9 (2004), <<http://library.cern.ch/HEPLW/9/papers/3/>>.

¹⁷ *SHERPA: securing a hybrid environment for research preservation and access*,

<<http://www.sherpa.ac.uk/index.html>>.

¹⁸ OAIS (Open Archival Information System) fornisce un modello funzionale per l'archiviazione dei dati e per la conservazione nel tempo dell'informazione digitale: OAIS, <http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html>; H. JAMES, et al., *Feasibility and requirements study on preservation of e-prints*, cit., p. 48-50; A. SALARELLI – A.M. TAMMARO, *La biblioteca digitale*, cit., p. 171-172; A. SCOLARI, et al., *Appunti per la definizione di un set di metadati gestionali-amministrativi e strutturali per le risorse digitali*, ICCU, 2001 <<http://www.iccu.sbn.it/MAG/metaAG1.pdf>>, (novembre 2004).

¹⁹ S. ANDERSON – S. PINEFIELD, *Preserving e-prints: scaling the preservation mountain*, intervento presentato al CURL/DPC Forum "Digital preservation in institutional repositories", cit.

²⁰ Esiste una lista di discussione degli utilizzatori di DSpace, dedicata proprio agli aspetti della preservazione: *DSpace-preservation*, <<http://mailman.mit.edu/mailman/listinfo/dspace-preservation>>.