

Accesso persistente per oggetti digitali ad accesso libero

Una delle principali finalità degli archivi ad accesso aperto (Open Access Archives) è di favorire una più ampia disseminazione della produzione scientifica e accademica. I docenti ed i ricercatori pubblicano on-line contenuti digitali (digital assets) ad accesso libero per diffondere i risultati delle proprie ricerche e per condividerli con altri studiosi ed esperti. Le università, i centri di ricerca, le biblioteche e – limitatamente a sottoinsiemi delle proprie collezioni - anche i musei, gli archivi ed altre istituzioni culturali hanno promosso negli ultimi anni la costituzione di archivi ad accesso aperto. Queste collezioni digitali avranno una importanza sempre maggiore nel costituire il patrimonio informativo, culturale e scientifico (information literacy) da tramandare alle prossime generazioni. In futuro, una parte consistente – probabilmente la più rilevante – della memoria delle istituzioni scientifiche, accademiche e culturali sarà formata da oggetti “nati” digitali, conservati in archivi ad accesso aperto. Per garantire la sopravvivenza di queste risorse e la loro citazione occorre che le comunità accademiche e scientifiche si impegnino fortemente nella conservazione a lungo termine dei propri archivi digitali.

Contesto e obiettivi

JISC ha definito la conservazione del digitale come “ the series of actions and interventions required to ensure continued and reliable access to authentic digital objects for as long as they are deemed to be of value”. In questa affermazione tre sono i concetti chiave da sottolineare nel contesto degli archivi ad accesso aperto (OAA):

- I repository degli OAA memorizzano oggetti digitali che prevalentemente corrispondono a contenuti che fin dall’origine sono digitali (born digital) e non hanno un equivalente di tipo analogico.
- L’attendibilità e la disponibilità dei contenuti degli OAA sono due qualità rilevanti per gli autori e per la comunità accademica e scientifica.
- L’accesso permanente ai contenuti degli OAA – ovvero la possibilità di continuare ad usare gli oggetti digitali nel tempo – è il requisito per consentire la citazione, la scoperta, la fornitura e l’uso delle risorse in modo durevole dopo la loro creazione ed il deposito negli archivi.

L’accesso deve essere garantito, fintanto che è ritenuto necessario, prevenendo e controllando le alterazioni ed i guasti dei supporti e dei sistemi, le anomalie di funzionamento dei repository e gli effetti dell’evoluzione tecnologica. A tal fine i processi per la conservazione a lungo termine devono essere basati su modelli organizzativi, politiche e strategie ben definite.

Modelli

L’OAIS Reference Model definisce un Open Archival Information Systems (OAIS) come “an archive, consisting of an organization of people and systems, that has accepted the responsibility to preserve information and make it available for a Designated Community”. Questo modello fornisce la specifica completa di un repository dal punto di vista funzionale ed informativo e stabilisce “mandatory responsibilities that an organization must discharge in order to operate an OAIS archive”.

L’applicazione dei principi definiti dall’OAIS ed in particolare l’implementazione di archivi ad Accesso Aperto in conformità ai modelli funzionale e informativo di OAIS sono un prerequisito chiave per la costruzione di repository affidabili (trusted) e per assicurare la conservazione a lungo termine dei loro contenuti digitali.

Modelli

La conservazione del digitale rafforza la fiducia degli autori e degli utenti negli archivi ad Accesso Aperto. Un “trusted repository” deve garantire almeno le seguenti proprietà:

- Autenticità. La certezza che una risorsa digitale appartenga a chi afferma di averlo creato. L’autenticità implica anche il “non ripudio”, cioè la certezza che chi ha creato l’oggetto non possa negare di esserne l’autore. La firma digitale e il digital watermarking sono tecniche usate per garantire l’autenticità degli oggetti digitali.

Ulteriori informazioni e risorse

CrossRef è un sistema per l'identificazione persistente di contenuti digitali in ambito accademico e per la gestione delle citazioni (link) ai full-text degli editori attraverso il DOI. URL: <http://www.crossref.org/02publishers/16openurl.html>

Citebase Search è un indice di citazioni semi autonomo per la letteratura libera on-line prodotta nell'ambito della ricerca. Effettua l'harvesting da archivi conformi a OAI-PMH. URL: <http://www.citebase.org/>

OAI-ORE Open Archives Initiative Object Reuse and Exchange. URL: <http://www.openarchives.org/ore/>

The Open Archival Information System (OAIS) Reference Model (standard ISO 14721). URL: <http://ssdoo.gsfc.nasa.gov/nost/isoas/>

DigitalPreservationEurope (DPE) una collezione di risorse on-line rilevanti e utili per studi e progetti sugli archivi digitali. URL: <http://www.digitalpreservationeurope.eu/resources/>

Digital Preservation Coalition (DPC) è stata costituita nel 2001 per coordinare una serie di azioni congiunte rivolte ad affrontare le sfide derivanti dalla conservazione delle risorse digitali in Inghilterra e per lavorare con altri partner di livello internazionale. URL: <http://www.dpconline.org>

JISC repositories and preservation programme ha per scopo lo sviluppo di un Information Environment volto a supportare i repository digitali e la conservazione del digitale, inclusi strumenti di cross-searching tra gli archivi. URL: http://www.jisc.ac.uk/whatwedo/programmes/programme_rep_pres.aspx

International Research on Permanent Authentic Records in Electronic Systems (InterPARES) due progetti (InterPARES 1 1999-2001 e InterPARES 2 2002-2006) che hanno elaborato conoscenze teoriche e metodologiche essenziali nella conservazione a lungo termine di registrazioni archivistiche create o conservate in forma digitale. URL: <http://www.interpares.org/>

PREMIS (PREservation Metadata: Implementation Strategies) Resources. PREMIS è un'iniziativa congiunta di OCLC e RLG volta a sviluppare un'insieme base di metadati per la conservazione, di larga applicazione nella comunità che si occupa di digital preservation, e a identificare e valutare strategie per la codifica, la memorizzazione e la gestione dei metadati nei sistemi di digital preservation. URL: <http://www.oclc.org/research/projects/pmwg/resources.htm>

The - Integrità. La capacità di mantenere la correttezza e la completezza dei dati prevenendo cambiamenti accidentali o dolosi (i.e. la corruzione dei dati). Il calcolo e il salvataggio dell'impronta digitale (un bit o byte checksum ricavato con algoritmi come MD5) sono una tecnica per poter verificare e scoprire a posteriori se qualche modifica ha alterato un oggetto digitale dopo la sua archiviazione iniziale.

- Affidabilità e Disponibilità. L'affidabilità si riferisce alla capacità delle componenti hardware e software di operare e comportarsi in base alle specifiche, senza anomalie o errori (in teoria completamente, in pratica garantendo una percentuale molto elevata). La Disponibilità è la percentuale di tempo nella quale un sistema, un'applicazione o un componente funzionano regolarmente rispetto al tempo totale di funzionamento richiesto. I salvataggi di dati e delle applicazioni, gli antivirus, i firewall, gli aggiornamenti dei sistemi operativi e dei software applicativi, le configurazioni hardware basate su componenti fault tolerant e ridondanti sono alcune delle tecniche comunemente adottate per garantire elevate percentuali di affidabilità e disponibilità.

- Possibilità di riuso successivo. La capacità di accedere ad una risorsa digitale fintanto che l'archivio e l'istituzione che la conservano decidono di supportarla. I contenuti digitali prodotti in ambito accademico e scientifico caratterizzati da un valore durevole devono poter essere cercati, trovati e (ri)utilizzati anche per molto tempo (per esempio attraverso l'uso di identificatori persistenti e il mantenimento dei formati e dei supporti fisici).

La conservazione del digitale deve comprendere tutti i dati dei repository che ospitano i contenuti digitali, non solo gli oggetti digitali veri e propri, ma anche i loro metadati, gli identificatori e i metadati dei resolver.

Politiche e strategie

- Sistemi di citazione affidabili basati su identificatori unici e persistenti: occorre fidarsi nell'integrità e nella autenticità delle risorse digitali citate, garantita dalle strategie di conservazione a lungo termine, e prevenire il fenomeno dei link interrotti (il ben noto errore "404 Not Found") adottando soluzioni tecniche ed organizzative che superino i limiti delle citazioni basate sugli URL.
- Devono essere definite appropriate politiche organizzative non solo per stabilire quali contenuti digitali conservare e per garantire il sostegno finanziario e organizzativo, ma anche per definire chi è responsabile della conservazione. I ruoli e le responsabilità devono essere definiti accuratamente, le conoscenze e le competenze devono essere costantemente sostenute e la formazione deve essere promossa con continuità.
- La conservazione del digitale richiede una conoscenza multidisciplinare e deve essere fondata sulla stretta collaborazione di professionisti ed esperti appartenenti a differenti aree.
- Occorre promuovere ed incrementare la consapevolezza degli autori nel seguire ben definite politiche e strategie di conservazione per garantire la continuità dell'accesso ai contenuti digitali da loro prodotti.
- I metadati contribuiscono ad assicurare la disponibilità e l'accesso persistente alle risorse digitali: la pianificazione della conservazione del digitale deve includere anche strategie e processi per la conservazione dei metadati.
- La revisione e la valutazione (auditing and assessment) costanti sono un impegno necessario per fare fronte ai cambiamenti di tipo organizzativo, delle politiche, legali e tecnici.