

Semplicemente DOI

La combinazione di standard di identificazione e tecnologie di risoluzione è ciò che contraddistingue la “nuova” generazione di identificatori persistenti

Paola Mazzucchi

Responsabile Ufficio editoriale mEDRA
paola.mazzucchi@aie.it

Identificare in modo univoco e persistente documenti, contenuti e risorse è esigenza da sempre sentita laddove vi sia la necessità di garantirne reperibilità e accesso tempestivi e a lungo termine, si pensi ad esempio all'ISBN per le pubblicazioni periodiche, all'ISSN per quelle periodiche o all'ISMN per la musica a stampa.

Con l'avvento massiccio delle nuove tecnologie digitali e con esse di nuovi modelli di messa a disposizione delle risorse, tale esigenza si è fatta sempre più pressante da un lato per facilitare la gestione interna di archivi sempre più complessi ed eterogenei, dall'altro per fornire al pubblico di potenziali fruitori di tali risorse – un pubblico sempre più delocalizzato dal punto di vita territoriale – strumenti di supporto alla ricerca e alla condivisione della conoscenza. In questo contesto fortemente mediato dalle tecnologie, e dalle tecnologie di rete in particolare, la necessità di identificare con precisione e in modo non ambiguo ciascuna singola risorsa si accompagna indissolubilmente alla necessità di fornirne un meccanismo di accesso e reperimento semplice e immediato, possibilmente esso stesso basato su tecnologie di rete. Per questo motivo si sono sviluppate soluzioni che alla funzione base di identificazione integrano la funzione di risoluzione, ovvero la possibilità di rimandare in modo immediato alla risorsa stessa o a maggiori informazioni sulla risorsa a partire dallo stru-

mento utilizzato per l'identificazione. La combinazione di standard di identificazione e tecnologie di risoluzione è ciò che contraddistingue la “nuova” generazione di identificatori persistenti.

Inutile dire che esistono molteplici soluzioni a disposizione, più o meno standard, più o meno diffuse, più o meno interoperabili; altrettanto inutile dire che esistono diverse “scuole di pensiero” ciascuna con le proprie predilezioni e le proprie idiosincrasie.

Nel panorama degli identificatori persistenti il DOI è una di queste possibili soluzioni. Vediamone le caratteristiche salienti, a partire da alcune delle più diffuse “false credenze” in materia.

Il DOI è un servizio commerciale.

Che il DOI non possa, o non debba, essere assegnato a risorse e pubblicazioni che non abbiano un aperto scopo commerciale è un fraintendimento piuttosto grossolano. È vero che il DOI System nasce per rispondere all'esigenza di un gruppo di editori internazionali, attivi commercialmente nel settore STM; tuttavia questa connotazione iniziale si è andata stemperando soprattutto grazie alla nascita e diffusione di modelli di business per le pubblicazioni scientifiche come l'Open Access che ha stimolato nuovi attori non commerciali ad entrare nel settore editoriale. Così oggi tra i membri dell'IDF (International DOI Foundation, <www.doi.org>) troviamo editori come Springer, Elsevier e Wi-

ley, ma anche il JISC (Joint Information System Committee, <www.jisc.ac.uk>) o il CENL (Conference of European National Librarians, <www.cenl.org>) e, ad esempio, nel 2009 oltre il 50% dei clienti dell'agenzia di registrazione operante anche in Italia (mEDRA, <www.medra.org>) è composto da dipartimenti universitari e centri di ricerca, la maggioranza dei quali assegna il DOI ad articoli di riviste Open Access.

Altrettanto vero è che il DOI si rivolge in prima istanza agli editori, ma il termine “editore” viene interpretato nella sua accezione più ampia per indicare qualsiasi soggetto (pubblico, privato, persona fisica o giuridica) che a vario titolo rende disponibile alla collettività, ovvero rende pubblico, o pubblica che dir si voglia, un contenuto, a prescindere dalla forma e dal supporto su cui tale contenuto venga pubblicato. Secondo questa definizione sono editori a cui il DOI si rivolge le case editrici in senso stretto, ma anche le university press, i dipartimenti universitari, i singoli autori che utilizzano servizi di autopubblicazione, i centri di ricerca e tutte le istituzioni pubbliche e private che abbiano una propria attività editoriale.

Il DOI è incompatibile con l'Open Access.

Che il DOI sia in qualche modo incompatibile con l'Open Access è uno dei più diffusi “falsi miti” in materia, il DOI infatti è completamente neutrale rispetto al modello di business della pubblicazione.

ne cui viene assegnato, anzi in un certo qual modo può essere un valido alleato dell'Open Access per garantire visibilità e rilevanza internazionale alle pubblicazioni. Il DOI è infatti ampiamente utilizzato a livello internazionale per identificare, citare e far circolare articoli e pubblicazioni dei ricercatori europei, sia attraverso servizi dedicati come il Cross Linking, sia nelle maggiori banche dati e repository; in questo modo pubblicazioni Open Access e pubblicazioni commerciali possono coesistere, essere ricercate e reperite parallelamente secondo meccanismi standard e condivisi. La corretta gestione dei metadati associati alla registrazione del DOI inoltre può supportare gli editori nella diffusione delle informazioni a soggetti terzi, come per esempio DOAJ (Directory of Open Access Journals, <www.doja.org>) o PMC (PubMed Central, <www.pubmedcentral.nih.gov>) attraverso le opportune trasformazioni di formati (ormai tutti basati su XML).

Il DOI costa troppo. Che il DOI abbia un costo è una verità incontrovertibile; che il DOI costi troppo invece è un altro mito difficile da eliminare e alimentato dal particolare modello di business adottato dalla maggioranza delle agenzie di registrazione: il singolo registrante paga per l'assegnazione del DOI e per i servizi a esso associati. Perché allora dover pagare quando tra le alternative a disposizione ci sono PURL, ARK, URN che sono gratuiti? La realtà è che nessuna delle alternative è completamente esente da costi per "l'utilizzatore finale"; ciò che varia sono le voci di costo che, in alcuni casi, appaiono meno evidenti, ma non per questo si debbono considerare inesistenti.

Nel modello del DOI System, il registrante (ovvero l'editore o un intermediario per suo conto) ha un costo legato al numero di DOI registrati, ma nessun costo in termini

di implementazione dei servizi associati al DOI.

Il servizio di risoluzione è infatti sviluppato, gestito e mantenuto centralmente dal CNRI, garantendo così robustezza, stabilità e performance del sistema a lungo termine

Il mantenimento dei metadati associati al DOI, tra cui l'associazione tra DOI e risoluzioni, è affidata alle agenzie di registrazione che per essere tali devono rispettare precisi criteri in materia di integrità dei dati e disponibilità dei servizi.

L'aggiornamento di software e hardware, così come l'implementazione di nuovi servizi è messo a fattore comune tra le diverse agenzie di registrazione che ne sostengono anche i costi attraverso la partecipazione all'infrastruttura sociale dell'IDF.

Il DOI è solo per oggetti digitali.

Che il DOI possa essere assegnato alle sole risorse digitali è un mito difficile da sfatare, nato a causa dell'interpretazione errata del significato dello svolgimento dell'acronimo DOI: Digital Object Identifier. L'aggettivo *digital* infatti non si riferisce tanto a *object* quanto a *identifier*, il DOI dunque sarebbe – correttamente – un identificatore digitale di oggetti e non un identificatore di oggetti digitali. La natura digitale dell'identificatore è così legata non all'oggetto che viene identificato, ovvero al referente, che può essere quindi analogico (cartaceo) o digitale, ma alla modalità di utilizzo dell'identificatore, ovvero alla possibilità di utilizzare (azionare) l'identificatore in rete grazie al processo di risoluzione. In altre parole, cliccare il DOI e farci qualcosa.

Infatti, la risoluzione può essere definita, in generale, come quel processo attraverso il quale si sottopone un identificatore a un servizio di rete e se ne riceve in risposta un insieme di informazioni a esso associate.

Un esempio di servizio di risoluzio-

I numeri del DOI

Primo DOI registrato ufficialmente nel 2000

Oltre 43 milioni di DOI registrati a fine 2009

Oltre 60 milioni di risoluzioni al mese

Oltre 4.000 registranti nel 2009

6 agenzie di registrazione nel mondo

Fonte: International DOI Foundation, 2010

ne è quello fornito dal DNS (*Domain Name System*): dato un nome di dominio (es. <www.doi.org>), il DNS restituisce l'indirizzo IP (es. 132.151.1.146) associato a esso, utilizzato dal computer client per instradare correttamente le richieste verso il computer server identificato dal nome di dominio. Tuttavia, a differenza del citato DNS, non vi sono vincoli particolari sui tipi di dato a cui un DOI può risolvere: possono essere URL che rappresentano ubicazioni in rete di particolari istanze dell'oggetto identificato dal DOI, servizi, come ad esempio le e-mail, o un insieme di metadati descrittivi dell'oggetto identificato. Diventa quindi essenziale la distinzione tra ciò che un DOI identifica e ciò a cui un DOI risolve: l'entità identificata da un DOI, detta oggetto di prima classe (*first-class object*), può essere una qualsiasi risorsa; le informazioni associate a un DOI e restituite dal servizio di risoluzione del DOI System sono invece attributi di tale risorsa, e non è detto che permettano di accedere a una sua manifestazione.

Per fare un esempio, la risoluzione di un DOI che identifica un articolo di rivista potrebbe restituire l'URL di una copia in formato PDF di quell'articolo, ma potrebbe anche restituire delle informazioni sul suo editore, l'indicazione delle piattaforme da cui l'articolo è scaricabile o informazioni su come abbonarsi alla rivista in formato cartaceo. La riso-

luzione del DOI inoltre può essere singola (ovvero un DOI una risoluzione, con un meccanismo simile al PURL) oppure multipla (ovvero un DOI più risoluzioni e più informazioni). Questa distinzione è analoga a quella tra l'ISBN di un libro, che lo identifica in modo univoco nel mondo editoriale, e le segnature, che indicano dove è possibile trovarne delle copie.

Il DOI può essere dunque assegnato a qualsiasi risorsa a prescindere dal formato o dal supporto in cui essa è pubblicata.

mEDRA: multilingual European DOI Registration Agency

Creata nel 2002 per iniziativa congiunta dell'Associazione italiana editori, che gestisce anche l'Agenzia ISBN italiana, e il Consorzio inte-

runiversitario di supercalcolo CINECA, mEDRA offre servizio di registrazione del DOI a un'ampia gamma di tipologie di editori in Italia e all'estero tra cui Germania, Austria, Spagna, Francia, Regno Unito, Paesi Bassi, Turchia. In Germania mEDRA ha una propria Agenzia locale che copre il mercato di lingua tedesca e che è gestita in collaborazione con l'Agenzia ISBN tedesca. mEDRA inoltre fornisce tutta l'infrastruttura tecnologica all'agenzia di registrazione DOI della Comunità Europea,

OPOCE, che assegna un DOI a tutti i documenti e le pubblicazioni dell'Europa, compreso l'*Official Journal*, la gazzetta ufficiale europea pubblicata nelle 27 lingue comunitarie. Nonostante la forte vocazione internazionale, mEDRA è e rimane un'azienda italiana, radicata nel territorio nazionale e in grado di offrire servizi, supporto e consulenza tagliati per le specifiche esigenze del mercato italiano, pur sempre aggiornati allo scenario internazionale.

Abstract

Persistent identification of a resource to allow it to be found, retrieved and used is a major issue in the digital environment. The DOI system provides an actionable, interoperable, persistent way to identify any resource by using a robust, scalable, ready to use infrastructure. This article will clear up some of the most common misunderstandings around the DOI, explaining how and why the DOI is a digital identifier of objects, suitable for any kind of publishers and business model neutral, it can support Open Access strategies and it has a cost but it's not expensive.