

Dig *Italia*

Numero 0 - **2005**

Rivista del digitale nei beni culturali

ICCU-ROMA

Identificatori persistenti per gli oggetti digitali

Mario Sebastiani

ICCU

Fa impressione constatare quanto sia cambiato rapidamente, a causa della rivoluzione elettronica, il mondo dell'informazione e della comunicazione scritta. In un convegno sui cataloghi delle biblioteche, svoltosi nel lontano 1987, presso la *University of California* di Los Angeles¹, uno dei relatori delineava uno scenario prossimo venturo che sicuramente all'epoca poteva sembrare a molti alquanto fantascientifico: «Nella misura in cui la tecnologia della trasmissione dell'informazioni andrà avanti crescendo, l'informazione gestita in *qualsiasi* modo da una biblioteca sarà disponibile alle altre in una qualche forma senza bisogno che vi sia un trasferimento fisico di oggetti fisici. C'è ancora una lunga strada per arrivare a questo, ma la direzione è abbastanza chiara: quello che una biblioteca possiede sarà virtualmente presente in ogni altra biblioteca. La stessa collezione sarà virtualmente disponibile a tutti e le differenze tra i cataloghi locali saranno confinate a differenze nella forma di disponibilità. La distinzione tra cataloghi locali e collettivi si dissolverà in una pratica insignificanza. Se vogliamo pensare proficuamente alla catalogazione descrittiva, allora dobbiamo pensare ad un futuro nel quale le copie effettivamente leggibili costituiranno solo una piccola frazione della collezione disponibile in un dato luogo e nella quale le copie virtuali saranno trattate alla pari con le copie reali».²

Oggi, possiamo affermare che lo scenario prospettato da questo intervento era impreciso per difetto, perchè confinato al mondo delle biblioteche. Internet, invece ha trasformato questa visione in una realtà concreta, non solo per gli utenti delle biblioteche, ma anche per chiunque, ovunque nel mondo, si colleghi alla rete, anche da casa propria. Naturalmente l'esplosione della comunicazione online ha comportato, oltre agli enormi indiscutibili vantaggi di cui tutti facciamo esperienza in vario modo, anche svariati problemi che solo ora cominciano a essere percepiti, sia pure a stento, dal pubblico non specialista. In questo senso, uno dei problemi principali è rappresentato dall'identificazione permanente degli oggetti digitali in rete, siano essi file di testo, musica, video, immagini. La consapevolezza di questo

¹ *Conference on the Conceptual Foundations of Descriptive Cataloging*, University of California, Los Angeles, 14-15 febbraio 1987. Atti in: Elaine Svenonius (a cura di), *The conceptual foundations of descriptive cataloging*, San Diego: Academic Press, 1989.

² Patrick Wilson, *The Second Objective*, in: *The conceptual foundations of descriptive cataloging* cit., p.7.

problema è nata soprattutto in relazione all'identificazione dei testi elettronici, ma riguarda comunque tutti gli oggetti digitali in genere.

Elaine Svenonius, professore di Library Information Science, presso la *University of California* di Los Angeles, ha formulato così, in una pubblicazione di pochi anni fa, la questione: «Si incontrano problemi con gli indirizzi di Internet quando, per recuperare un dato documento, sono disponibili molteplici siti e mirror. Possono verificarsi problemi anche quando i siti per il recupero non sono distinti dai siti di pubblicazione. Ma il problema più grave capita quando cambiano gli indirizzi. Questo problema riguarda in particolare gli URL. L'URL di un documento può cambiare quando vengono fatti dei cambiamenti per riconfigurare l'hardware o riorganizzare il sistema. L'OCLC ha affrontato questo problema istituendo il servizio *Persistent Uniform Resource Locator (PURL)* che registra i cambiamenti degli indirizzi in modo da fornire agli utenti l'ultimo indirizzo di un documento. Ma questa soluzione è solo una misura temporanea».³

In altri termini, il punto è che per gestire e rendere effettivamente fruibili gli oggetti digitali distribuiti in rete e anche per garantire la loro conservazione duratura, è necessario disporre di sistemi che consentano di identificare in maniera chiara, univoca, non ambigua, ma soprattutto stabile nel tempo, questi oggetti, alla stessa stregua di quanto avviene, ormai da molto tempo, in modo semplice ed efficiente, per i documenti stampati (libri, periodici).

Identificare opere a stampa e testi elettronici: le differenze

Per illustrare le differenze tra opere a stampa e testi elettronici, relativamente al tema dell'identificabilità permanente, il modo più semplice di procedere è quello di descrivere un esempio concreto di ricerca. Nel 1813, Jane Austen descriveva, in *Orgoglio e pregiudizio*, la vita e il contesto sociale dell'aristocrazia di campagna inglese del secolo scorso e ne analizzava con delicata ironia i complessi rapporti tra il carattere individuale delle persone e la loro posizione sociale.⁴ Volendo leggere una copia originale inglese del testo di questa celebre opera, *Pride and prejudice*, potremmo percorrere due strade: la prima, quella più tradizionale, consisterebbe nel cercare una biblioteca che possieda l'opera a stampa. Potremmo allora procedere in tal senso, interrogando il catalogo online del "Servizio Bibliotecario Nazionale" (SBN).⁵ Il sistema, in risposta alla nostra richiesta, fornirà un elenco di edizioni che saranno identificate in primo luogo dai relativi dati bibliografici e, quando disponibile, anche da un codice sintetico, l'*International Standard Book Number* (ISBN). Questo codice viene pubblicato sull'ultima pagina

³ Elaine Svenonius, *The intellectual foundation of information organization*, Cambridge (Mass.): MIT Press, 2000, p.156.

⁴ La scrittrice inglese Jane Austen nacque a Steventon, Hampshire, nel 1775, e morì a Winchester nel 1817.

⁵ Cfr. <http://opac.sbn.it/>.

di copertina di un libro e consente di identificare immediatamente e inequivocabilmente quel dato libro. Nell'immagine 1 possiamo vedere la scheda bibliografica di un'edizione dell'opera di J. Austen, insieme al relativo codice ISBN, tratta dal catalogo online dell'SBN.

1.	Pride and prejudice / Jane Austen ; edited with an introduction by Luciana Pire. - Firenze : Giunti, 2001
2.	ISBN-88-09-02074-X.

1 - Un'edizione a stampa di Jane Austen identificata: 1. tramite i dati bibliografici; 2. tramite il codice ISBN.

Il codice ISBN è nato nel 1967, in Inghilterra, e nel 1970 è diventato uno standard ISO.⁶ Sebbene l'ISBN sia utile soprattutto a editori e librai, anche la gente comune può trovarsi in situazioni nelle quali questo codice si rivela una risorsa molto utile. Ad esempio, all'inizio di ogni nuovo anno scolastico, costituisce un'esperienza comune a molti studenti e a molti genitori, quella di controllare l'esattezza dei testi scolastici acquistati confrontando il codice ISBN presente sui libri con quello riportato sulla lista fornita dalla scuola.

Naturalmente la ricerca del testo non termina con l'identificazione di un'edizione: occorre anche accedere materialmente a un esemplare fisico di quella edizione. A questo scopo il sistema SBN fornisce, per ogni scheda relativa a un'edizione, un altro codice molto importante: il codice di *localizzazione* del libro. Questo è un numero che identifica quella o quelle biblioteche che, tra le tante che aderiscono all'SBN, possiedono una copia del libro in questione. Nel caso dell'edizione identificata dal numero "ISBN-88-09-02074-X", uno dei codici di localizzazione forniti dall'indice dell'SBN sarà "RM0267", vale a dire il codice corrispondente alla Biblioteca Nazionale Centrale di Roma (BNCR). Disponendo di questa informazione, potremo allora recarci presso la BNCR e chiedere di leggere il libro, tramite il prestito o la consultazione in sede. Non prima però, è ovvio, di esserci procurati un ultimo dato cruciale: la *collocazione* del testo nella biblioteca, vale a dire il codice che indica su quale scaffale, di quale magazzino della biblioteca, è fisicamente collocato il libro. Anche questo dato si può reperire sul web: basta interrogare il catalogo online della BNCR.⁷ Nel caso del nostro esempio, la collocazione sarà allora "AZD AST00 10372". Possiamo vedere i tre codici riuniti nella tab.2.

⁶ Come ISO 2108. L'amministrazione e il coordinamento internazionale del sistema ISBN sono svolti dall'Agenzia Internazionale ISBN con sede a Berlino presso la Biblioteca nazionale; cfr. <http://www.isbn-international.org>. Il ruolo di agenzia titolare dell'ISBN per l'area di lingua italiana è svolto dall'Associazione Italiana Editori (AIE); cfr. <http://www.isbn.it>. Per i periodici esiste l'International Standard Serial Number (ISSN); cfr. <http://www.issn.org:8080/pub/>.

⁷ <http://opac.bnrcr.metavista.it/>.

1.	codice ISBN:	ISBN-88-09-02074-X
2.	cod.di localizzazione:	RM0267
3.	codice di collocazione:	AZD AST00 10372

2 – Codici per l'identificazione, la localizzazione, la collocazione di un libro di J. Austen.

Da notare, per quanto riguarda questi dati, che un cambiamento dei codici di localizzazione e di collocazione, non pregiudica l'identificazione permanente dell'edizione. Anche se l'esemplare fisico in questione viene spostato da un punto all'altro della biblioteca, e quindi cambia la collocazione, l'esemplare continua a essere identificato dallo stesso codice ISBN. Lo stesso vale se cambia la localizzazione.

Nel caso dei testi elettronici su Internet, invece, le cose procedono alquanto diversamente. Come è noto sul *World Wide Web* (WWW) vi sono numerosi siti che rendono liberamente scaricabili le versioni elettroniche di opere a stampa che, per il tempo trascorso o altri motivi, non sono più protette dal diritto d'autore. Un sito del genere è quello denominato *Project Gutenberg*, forse la più "antica" biblioteca digitale della storia.⁸ Possiamo allora proseguire il nostro esempio, cercando in questo sito una eventuale versione elettronica dell'opera di J. Austen. Nell'immagine 3 è riportato l'indirizzo web di una versione elettronica dell'opera di J. Austen ottenuta interrogando il catalogo online del *Project Gutenberg*. Il testo elettronico è in formato PDF (*Portable Document Format*).⁹

<http://www.gutenberg.org/dirs/etext98/pandp12p.pdf>

3 – Indirizzo web che identifica e localizza, sul sito del Project Gutenberg, una versione elettronica dell'opera di J. Austen.

È evidente, a questo punto, che tra testi a stampa e testi elettronici vi è una differenza fondamentale: nel caso delle opere a stampa, tre distinti codici provvedono alle funzioni di identificazione, localizzazione e collocazione; nel caso invece dei testi elettronici, un unico codice assolve simultaneamente a tutte le funzioni. Col risultato che se l'indirizzo web del testo elettronico dovesse cambiare – e questo può avvenire per una quantità di motivi di ordine tecnico e/o amministrativo – allora qualsiasi citazione di quel testo smetterebbe di avere valore. Cliccando sul link che rinvia a quell'indirizzo web, otterremmo in risposta il messaggio di errore: «HTTP 404 -File non trovato». In assenza di ulteriori informazioni, non saremmo più in grado nemmeno di sapere di che testo si tratta.

⁸ *Project Gutenberg* è un progetto avviato nel 1971 da Michael Hart con l'obiettivo di costituire una biblioteca di versioni elettroniche liberamente riproducibili di libri stampati. Al settembre 2005 questa biblioteca digitale contava oltre 16.000 testi. Per il catalogo online cfr. <http://www.gutenberg.org/catalog/>.

⁹ Da notare che sul *Project Gutenberg* non viene riportata l'indicazione di edizione dei testi.

Internet si basa sul protocollo TCP/IP e gli indirizzi IP

Perché gli indirizzi web non garantiscono un'identificazione permanente degli oggetti digitali in rete? Per rispondere a questa domanda è necessario richiamare alcune nozioni fondamentali relative alla storia e al funzionamento di Internet. Il fondamento di questa rete è costituito dall'interconnessione globale di centinaia di migliaia di computer altrimenti indipendenti. Ciò che rende questa interconnessione possibile è l'uso di un insieme di standard di comunicazione, procedure e formati comuni in tutto il network, chiamati *protocolli*. A questo riguardo la *suite* fondamentale di protocolli di Internet è quella denominata TCP/IP (*Transport Control Protocol / Internet Protocol*). Il TCP/IP è un protocollo di interconnessione cosiddetto a pacchetti. O per meglio dire, di *commutazione* a pacchetto. La commutazione a pacchetto si differenzia dalla tradizionale commutazione a circuito, quella della rete telefonica tradizionale per intenderci, per il fatto che tra il mittente e il destinatario dei dati, non si determina una connessione stabile, ma i dati trovano volta a volta, sulla rete complessiva, il percorso più rapido per giungere a destinazione. Questa particolare modalità di interconnessione ha finito per prevalere perché più adatta a gestire il modo particolare in cui i computer si trasmettono i dati l'uno all'altro. I computer, infatti, quando comunicano tra di loro, lanciano brevi ma intense raffiche di dati, per poi rimanere silenziosi in attesa del flusso successivo. Nella commutazione a pacchetto questi flussi vengono trasmessi appunto come pacchetti di dati. Possiamo immaginare allora questi pacchetti che, come delle cartoline, vengono inoltrati da un computer a un altro finché non giungono a destinazione. In una rete di computer, quelli deputati unicamente a inoltrare e smistare i pacchetti di dati, sono chiamati *router*. L'insieme dei router e dei collegamenti tra loro costituiscono l'infrastruttura di base della rete Internet. L'enorme sviluppo di questa rete è dovuto proprio alla grande efficienza con cui questa infrastruttura, grazie alla interconnessione a pacchetti, è in grado di operare. Ritornando all'analogia delle cartoline, queste possono andare perse; possono essere inoltrate in modo disordinato e possono essere rallentate lungo il loro percorso da vari fattori. Lo stesso vale per i pacchetti Internet. Vi è quindi uno strato di rete che provvede a inoltrare i pacchetti, l'*Internet Protocol* (IP), e uno strato superiore, il *Transport Control Protocol* (TCP), che provvede a gestire gli inconvenienti che possono accadere: ad esempio rispedisce i pacchetti che si sono persi lungo il percorso o rimette in ordine quelli arrivati in disordine.

Naturalmente, affinché i pacchetti possano giungere a destinazione – e ricevere una risposta – è necessario che trasportino con loro, proprio come una cartolina postale, l'indirizzo del destinatario e quello del mittente. A questo scopo – in una rete TCP/IP – ogni computer è "targato" con uno specifico numero di riconoscimento, diverso da tutti gli altri. Questo numero si chiama *indirizzo IP*. È con questo numero che, nei pacchetti TCP/IP, vengono compilati il campo "destinatario del pacchetto" e "mittente del pacchetto". Gli indirizzi IP sono costituiti da quattro

gruppi di cifre divisi da un punto. Ogni gruppo è composto da una, due o tre cifre. Ad esempio 193.206.221.29 è un indirizzo Internet valido.

Questo sistema ha dato origine a una rete efficiente e affidabile che può essere ampliata a piacere, tanto che, pur avendo raggiunto una dimensione mondiale, le sue prestazioni continuano a crescere. Come hanno detto giustamente Robert Kahn e Vinton Cerf, gli inventori del protocollo TCP/IP: «Senza commutazione a pacchetto, non esisterebbe Internet così come la conosciamo».¹⁰

Il name-system di Internet è il Domain Name System (DNS)

Nella rete Internet, sopra lo strato di trasporto (rappresentato dal protocollo TCP/IP) operano altri protocolli che, usati insieme a determinati linguaggi per codificare le informazioni, consentono alla rete di erogare quei servizi per i quali è diventata così popolare. In primo luogo il protocollo HTTP (*Hyper Text Transfer Protocol*) e il linguaggio HTML (*HyperText Markup Language*). Insieme, queste due componenti costituiscono l'infrastruttura della rete ipertestuale mondiale del WWW.

Quando clicchiamo un link in una pagina web, il browser web non fa altro che inviare, mediante il protocollo HTTP, al computer remoto, corrispondente all'indirizzo web riportato nel link, una richiesta di visualizzazione della corrispondente pagina web. La pagina richiesta giungerà, codificata appunto in formato HTML, al nostro browser web che provvederà a visualizzarla sullo schermo del nostro pc. Lo stesso avviene se accediamo alla pagina web remota semplicemente digitando l'indirizzo web della pagina, nell'apposita barra degli indirizzi del browser. In tal caso, l'indirizzo sarà costituito dall'etichetta "http://" (che indica appunto il protocollo di alto livello che vogliamo usare) seguito dall'indirizzo web vero e proprio della pagina che ci interessa.

Sorge però, a questo punto, un interrogativo: se alle pagine web si accede con un indirizzo composto di lettere, che fine hanno fatto gli indirizzi numerici IP che abbiamo visto prima? Non sono scomparsi; in realtà l'indirizzo web è una modalità di accesso alle pagine web complementare all'accesso tramite l'indirizzo IP. Nell'immagine 4 possiamo vedere un esempio in proposito. Si può accedere alle pagine web, in genere, o con l'indirizzo IP o con l'indirizzo web.

1.	<code>http:// 193.206.221.29/</code>
2.	<code>http://opac.sbn.it/</code>

4 – Accesso al catalogo online dell'SBN digitando: 1. l'indirizzo IP; 2. l'indirizzo web.

¹⁰ Robert E. Kahn – Vinton G. Cerf, *What Is The Internet (And What Makes It Work)*, dicembre 1999, http://www.cnri.reston.va.us/what_is_internet.html.

Il sistema sul quale si basano gli indirizzi web tradizionali è il *Domain Name System* (DNS). Fin dalla fase sperimentale di Internet, che va dagli anni '70 ai primi anni '80, ci si rese conto che molte applicazioni di rete facevano riferimento al nome del computer (ad es. UCLA, USC-IS) piuttosto che all'indirizzo IP. Occorreva perciò creare una corrispondenza, all'interno della rete, tra gli indirizzi IP e i nomi dei computer. Una lista di questo genere, chiamata "host.txt" file (poiché era un semplice file di testo), venne mantenuta per un certo tempo da un gruppo denominato *Network Information Center* (NIC)¹¹. I computer collegati a Internet semplicemente scaricavano di tanto in tanto questa lista dal computer del NIC e facevano riferimento a essa per determinare quale indirizzo IP fosse associato a un determinato nome di computer. Tuttavia Internet crebbe così rapidamente che l'aggiornamento di tale lista divenne un compito sempre più arduo per il NIC.

Prevedendo che, con l'espansione della rete, questo problema sarebbe andato sempre più peggiorando, alcuni ricercatori cominciarono a impegnarsi per disegnare un sistema che permettesse di fornire le medesime informazioni in modo più distribuito.¹² Il risultato finale fu appunto il DNS che consente a centinaia di migliaia di *name server* di mantenere una piccola porzione del database globale contenente le informazioni che consentono di associare gli indirizzi IP con i nomi dei computer collegati a Internet.

Il sistema DNS è essenzialmente gerarchico: al vertice della gerarchia troviamo i *Top Level Domain name* (TLD) o i *country code Top Level Domain* (ccTLD). Ad esempio nell'indirizzo *iccu.sbn.it*, il sottodominio *iccu* appartiene al dominio *sbn* il quale a sua volta appartiene al ccTLD *it*.¹³ Come ha scritto Paul Mockapetris nel 1983, in occasione della pubblicazione dei primi standard DNS: «L'obiettivo dei nomi di dominio è di fornire un meccanismo per denominare le risorse in modo che i nomi possano essere utilizzati in differenti host, network, famiglie di protocolli, internet e organizzazioni amministrative».¹⁴

Quando un Host ha la necessità di comunicare con un altro Host, comunica al suo server DNS di riferimento il nome dell'Host remoto al quale vuole connettersi chie-

¹¹ Il gruppo risiedeva presso SRI International di Menlo Park (CA).

¹² Questo gruppo di ricerca era attivo presso lo Information Sciences Institute della University of Southern California.

¹³ Nel complesso sistema amministrativo di Internet, la gestione dei TLD è affidata a varie agenzie; ad es. il TLD.com è gestito da VeriSign Global Registry Services. La gestione dei ccTLD invece è affidata ai Regional Internet Registry (RIR). Il RIR della regione europea è il Réseaux IP Européens (RIPE). Al RIPE fanno capo le varie agenzie nazionali: quella italiana (responsabile del ccTLD.it) è rappresentata dall'Istituto di Informatica e Telematica del Consiglio Nazionale delle Ricerche (Pisa). Tutte queste organizzazioni fanno capo alla Internet Assigned Numbers Authority (IANA) la quale a sua volta è controllata dall'Internet Corporation for Assigned Names and Numbers (ICANN), una associazione no-profit fondata il 18 settembre 1998 in California, ma che opera sotto il diretto controllo del Dipartimento al Commercio degli Stati Uniti.

dendo di effettuare la *risoluzione* di tale nome nel relativo indirizzo IP. Se il server interrogato trova l'indirizzo ricercato all'interno del suo database, lo comunica all'Host richiedente che lo utilizzerà per connettersi all'Host remoto. Se il server non trova l'indirizzo, avvia una richiesta *ricorsiva*: sottopone cioè la medesima richiesta a un altro server DNS il quale a sua volta effettuerà la medesima ricerca. Nel caso che anche questa ricerca abbia esito negativo, la procedura andrà avanti finché la domanda non arriverà a un qualche server DNS che dispone dell'indirizzo ricercato: questo allora verrà finalmente ritrasmesso indietro fino all'Host richiedente.

Il WWW si basa su file di testo collegati tra loro mediante link ipertestuali basati appunto sul DNS. Altre applicazioni però, come la posta elettronica, si basano sul DNS: l'espressione che segue la chiocciolina (@) in un indirizzo di posta elettronica, è appunto un nome di dominio. È evidente allora l'importanza che questo *name-system* riveste all'interno dell'organizzazione complessiva della rete Internet. Il DNS però ha un limite apparentemente insuperabile: unisce in una unità praticamente indissolubile tanto la funzione di identificazione degli oggetti digitali quanto quella di localizzazione degli stessi sulla rete. Con il risultato che l'indirizzo web di un oggetto digitale non può cambiare senza che vengano meno, simultaneamente, tanto la localizzazione che l'identificazione dell'oggetto.

Gli standard URI, URL, URN

Agli sviluppatori di Internet apparvero presto chiari tanto i vantaggi del DNS quanto i suoi limiti. Si avviò quindi presto un dibattito, volto a definire quali fossero le modalità auspicabili di identificazione e localizzazione degli oggetti digitali in rete e a formalizzare queste modalità in standard opportuni. Il primo di questi fu lo *Uniform Resource Locator* (URL). Questo standard è stato definito per la prima volta nel 1994 da T. Berners-Lee, ideatore del WWW. Dalla definizione che ne ha fatto il suo autore, emerge come la commistione tra localizzazione e identificazione sia espressamente ricercata: «Gli URL vengono utilizzati per *localizzare* una risorsa fornendo una identificazione astratta della localizzazione della risorsa. Dopo aver localizzato una risorsa, un sistema può effettuare varie operazioni sulla risorsa che possono essere caratterizzate con parole quali *accesso*, *aggiornamento*, *sosti-*

¹⁴ Paul Mockapetris, *Domain names: Implementation and Specification, RFC 883*, November 1983, pag.ii. Cfr. <http://www.ietf.org/rfc/rfc883.txt>. L'attività di standardizzazione in Internet viene portata avanti principalmente da due organismi: l'*Internet Engineering Task Force* (IETF) e l'*Internet Architecture Board* (IAB). Entrambi fanno capo alla *Internet Society* (ISOC), un organismo internazionale, fondato nel 1992, del quale fanno parte più di 100 organizzazioni appartenenti a oltre 180 paesi. La sigla RFC (*Request For Comment*), con la quale l'ISOC pubblica i suoi standard, indica, a dispetto del nome, il documento definitivo. La sigla è un retaggio di quando gli organismi di standardizzazione di Internet erano strutture estremamente informali.

tuzione, ricerca di attributi. In generale per uno schema URL è necessario specificare solo il metodo di accesso». ¹⁵ Tutti i tradizionali indirizzi web, basati sul DNS, sono in pratica degli URL.

Ma la commistione tra identificazione e localizzazione, che sta alla base di questo standard, è estremamente insoddisfacente. Ristrutturazioni tecniche e amministrative rendono inevitabile cambiare, prima o poi, i nomi di dominio dei server sui quali sono allocati i testi digitali. Inoltre può insorgere la necessità di trasferire i testi da un server a un altro con un altro nome di dominio. In tutti questi casi, nel momento in cui si cambia l'URL, i documenti vengono di fatto ridenominati.

L'intrinseca non-persistenza dello standard URL ha portato la comunità degli sviluppatori di Internet a definire due nuovi standard più avanzati: lo *Uniform Resource Identifier* (URI) e lo *Uniform Resource Name* (URN). Secondo un primo orientamento, questi due standard dovevano costituire l'espressione di una netta demarcazione tra identificazione e localizzazione. Gli identificatori sarebbero dovuti ricadere in due classi: quella degli identificatori che specificano la localizzazione di una risorsa (URL) e quella degli identificatori che specificano il nome della risorsa a prescindere dalla localizzazione (URN). Tutti e due avrebbero rappresentato una particolare, differente, concretizzazione del più generico standard URI; infatti, come leggiamo in un documento del 2002: «Senza perdere generalità, è ragionevole affermare che lo spazio URI era stato concepito per essere partizionato in due nuove classi: URL e URN. Così, ad esempio, «HTTP:» costituiva uno schema URL, mentre «ISBN:» sarebbe diventato (prima o poi) uno schema URN. Ogni nuovo schema sarebbe dovuto ricadere in una di queste due classi». ¹⁶

Ma, all'inizio del 2005, il principale oppositore dell'approccio dicotomico, T. Berners-Lee, è tornato alla carica ribadendo che, sulla rete, non è possibile separare tra loro le funzionalità di localizzazione e di identificazione; gli standard quindi devono limitarsi a prendere atto di questa realtà: «Uno schema individuale non deve necessariamente essere classificato o come "nome" o come "localizzatore". In un qualsiasi schema dato vi possono essere dei casi di URI che possiedono le caratteristiche di nome, di localizzatore o di entrambi, spesso in dipendenza dalla perseveranza e dalla cura della *naming authority* nell'assegnare gli identificatori, piuttosto che dalla qualità dello schema. Perciò le specificazioni future, insieme alla documentazione correlata, dovranno usare il termine più generale URI, piuttosto che i termini più restrittivi URL e URN». ¹⁷

¹⁵ Tim Berners-Lee – Larry Masinter – Mark McCahill (Eds), *Uniform Resource Locators* (URL), RFC 1738. Dicembre 1994, <http://www.ietf.org/rfc/rfc1738.txt>.

¹⁶ Michael Mealling – Ray Denenberg (Eds.), *Report from the Joint W3C/IETF URI Planning Interest Group: Uniform resource Identifiers (URIs), URLs., and Uniform Resource Names (URNs): Clarifications and Recommendations*, RFC 3305. August 2002, pag.2, <http://www.ietf.org/rfc/rfc3305.txt>.

¹⁷ Tim Berners-Lee – Larry Masinter – Mark McCahill (Eds), *Uniform Resource Locators* (URL), RFC 1738 cit.

Ovviamente, la replica più ovvia a una posizione di questo genere, è che: «Gli URL, comunemente intesi, sono dimostrabilmente non persistenti; ridefinirli come URI non risolve niente».¹⁸ A questo si può aggiungere che nemmeno risolve qualcosa il fatto di ridefinire gli URL come URN poiché, date le caratteristiche di questo standard: «La mancanza di un qualsiasi supporto infrastrutturale richiederà che ogni implementazione di URN dovrà sviluppare il suo proprio sistema di risoluzione, sottoforma di *plug-in* o come *proxy-server*».¹⁹

Perciò, alla luce di queste brevi considerazioni, su una materia che in realtà richiederebbe molto più spazio, possiamo dire che, nell'approfondire il tema degli identificatori persistenti, converrà limitarsi a esaminare le esperienze concrete realizzate finora, tralasciando di considerare se e quanto queste esperienze si conformino a uno standard piuttosto che a un altro. L'unico punto fermo è che gli indirizzi web tradizionali, basati sul sistema DNS, sono senza dubbio degli URL. Ma, per quanto riguarda gli URI e gli URN, sembra si possa condividere il punto di vista di N. Paskin, direttore dell'*International Doi Foundation*, quando afferma che: «Attualmente non esistono implementazioni pratiche sufficientemente diffuse di queste specificazioni che operino quali sistemi di denominazione degli oggetti: sia l'URI che l'URN sono specificazioni, non implementazioni operanti di per sé».²⁰

Identificatori basati sul DNS: 1) il Persistent Uniform Locator (PURL)

Possiamo allora tentare di classificare i sistemi di identificazioni persistente proposti finora, non in base allo standard cui si conformano, ma piuttosto secondo il sistema di risoluzione (cioè il sistema di traduzione degli indirizzi in numeri IP) che utilizzano. Tra gli identificatori permanenti che si basano sul tradizionale DNS, il più noto è senz'altro il PURL (*Persitent Uniform Resource Locator*) sviluppato dall'OCLC (*Online Computer Library Center*). Nella sostanza, il PURL è un tradizionale indirizzo web. Quello che il PURL offre in più, rispetto a un comune URL, è un servizio di reindirizzamento. Questo servizio consente, quando la localizzazione web di una risorsa elettronica cambia, di deviare sul nuovo indirizzo web tutte le richieste che continuano a pervenire al vecchio indirizzo. Funzionalmente un PURL è un URL che punta a un servizio intermedio di risoluzione. Il PURL, in risposta alle richieste del browser, invia un normale URL che può poi essere utilizzato per accedere alla risorsa remota in modo tradizionale. Nel linguaggio web questo è un comune servizio di *redirect*. Qui non c'è niente di nuovo: il *redirect* è un servizio già previsto dallo standard HTTP e comunemente usato. Quando si cambia un URL, sul

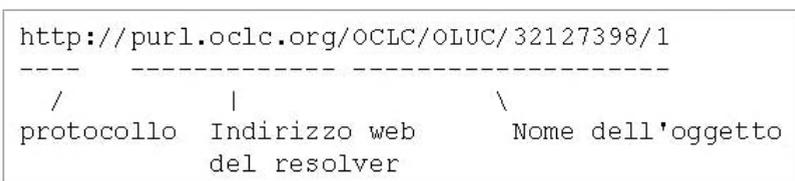
¹⁸ Norman Paskin (Ed.), *The DOI Handbook*. Edition 4.2.0, International Doi Foundation, 2005, p. 35, <http://dx.doi.org/10.1000/186>.

¹⁹ *The DOI Handbook* cit., p. 36. In genere con l'espressione proxy-server si indica una macchina che consente connessioni indirette da un servizio a un altro su un network.

²⁰ *The DOI Handbook* cit., p.36.

sito del vecchio URL si può implementare il servizio *redirect*, secondo le modalità previste dagli standard HTTP e HTML, in modo da reindirizzare le richieste mandate al vecchio URL, verso quello nuovo. Il PURL, a differenza di un comune indirizzo web, si limita a usare in maniera sistematica il servizio *redirect* al fine di garantire la persistenza dell'identificatore (o almeno per fornire un servizio che somigli quanto più possibile a un servizio di identificazione persistente). L'URL associata al PURL può cambiare, il PURL no. La persistenza offerta dal PURL sta tutta qui e i PURL sono a tutti gli effetti degli URL.

A differenza del comune URL, il PURL è composto di tre parti: (1) l'indicatore di protocollo, (2) l'indirizzo del resolver e (3) il nome. Nell'immagine 5 possiamo vedere come si articolano queste tre componenti. Da notare che l'indirizzo del resolver è rappresentato dal nome di dominio (ovvero dall'indirizzo IP) del resolver PURL: questa parte del PURL continua ad essere risolta dal tradizionale DNS.



5 – Il Persistent Uniform Resource Locator (PURL) dell'OCLC.

Creare un PURL non è difficile. Occorre in primo luogo essere un utente registrato di un resolver PURL: ad esempio quello dell'OCLC.²¹ Ci si registra, alla stessa stregua di tanti altri siti web, immettendo nella form opportuna i propri dati e creando un proprio identificativo e una password.²² Un utente registrato può quindi creare i suoi propri PURL purchè rispetti determinate condizioni: in primo luogo che esista il dominio top-level dell'indirizzo web che si vuol rendere persistente. La parte del PURL che costituisce il nome della risorsa viene decisa dal creatore del PURL. La componente relativa all'indirizzo del resolver, no. I nomi possono essere arbitrari e non è necessario che vi sia alcuna correlazione tra il nome all'interno del PURL e l'URL associato con esso. Gli utenti possono assegnare un PURL a ogni risorsa discreta per la quale si desidera un accesso affidabile nel tempo. Ad esempio una home page, un giornale elettronico, un articolo individuale, uno scritto, sono tutti oggetti digitali suscettibili di ricevere un identificatore PURL. Invece le risorse che non sono discrete, come le sezioni di un documento, i grafici o quegli schemi che non avrebbero senso al di fuori del documento che li contiene, non sono adatti a essere identificati tramite un PURL.

²¹ Cfr. <http://purl.oclc.org/>.

²² La form del resolver PURL dell'OCLC si trova su <http://purl.oclc.org/maint/register.html>.

Per creare un PURL basta seguire la procedura guidata, dopo essersi registrati, presente sul resolver PURL. I PURL non vengono aggiornati automaticamente quando l'URL associato a essi varia. L'aggiornamento delle informazioni presenti nel server PURL è un compito di cui deve farsi carico un *maintainer*. La manutenzione del PURL viene svolta connettendosi al resolver PURL e quindi compilando l'opportuna *form* di manutenzione. Solo i *maintainer* autorizzati possono modificare un PURL. Se un *maintainer* dovesse eliminare il PURL, il resolver, invece di fornire una destinazione, fornirebbe la storia di quel PURL (gli indirizzi associati ad esso fino all'ultimo giorno in cui era operativo)²³.

1.	<code>http://purl.oclc.org/keith/home</code>
2.	<code>http://home.columbus.rr.com/shafer/</code>

6 – Un PURL (1.) e l'URL associato (2.).

Gli utenti possono selezionare (cliccandoci sopra) un PURL su una pagina web o in un documento e il PURL verrà risolto verso l'URL associato che il browser userà poi per accedere alla risorsa. Gli utenti possono inserire i PURL nelle pagine web, nei documenti o in altre risorse con la ragionevole sicurezza che il PURL rimarrà persistente nel tempo. I link rimarranno validi anche se l'URL associato dovesse cambiare. Questo non significa che un PURL cambia magicamente il suo proprio URL associato quando la risorsa indirizzata viene spostata. È compito dei *maintainer* del PURL provvedere a ciò.

Da sottolineare che il PURL, a differenza dell'ISBN, è un identificatore *azionabile*: cioè, se viene inserito in un documento elettronico, cliccandoci sopra si recupera il documento identificato.

Identificatori basati sul DNS: 2) l'Archival Resource Key (ARK)

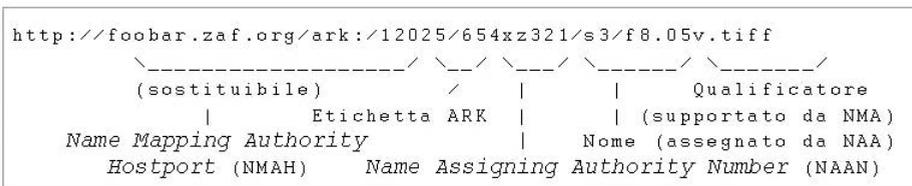
L'ARK, *Archival Resource Key*, è uno schema di identificazione per l'accesso persistente agli oggetti digitali (includere immagini, testi, data set, e aiuti alla ricerca) che è attualmente in corso di test e che viene implementato dalla *California Digital Library* (CDL) per le proprie collezioni digitali.²⁴ L'origine dell'ARK la si deve alla *National Library of Medicine*. L'ARK è l'unico indicatore che antepone gli aspetti procedurali del servizio di identificazione permanente, agli altri aspetti più prettamente tecnologici: «Un principio fondamentale dell'ARK è che la persistenza è una pura materia di servizio e che non è inerente all'oggetto né viene conferita a questo da una particolare sintassi di identificazione. Il massimo che un identificatore può fare è di ricondurre gli utenti a questi servizi».²⁵

²³ Cfr. <http://purl.oclc.org/docs/inet96.html>.

²⁴ Cfr. <http://www.cdlib.org/inside/diglib/ark/>.

²⁵ John A. Kunze, *Towards Electronic Persistence Using ARK Identifiers*, University of California, July 2005, <http://www.cdlib.org/inside/diglib/ark/arkcdl.pdf>.

In concreto l'ARK è un normale URL, caratterizzato da una forma particolare, che gli conferisce alcune delle condizioni necessarie per l'identificazione persistente. Un ARK è rappresentato da una sequenza di caratteri che contiene l'etichetta «ark:» e che, opzionalmente, può essere preceduta dalla parte iniziale di un URL. Questo URL, detto *Name Mapping Authority Hostport* (NMAH) è mutevole e sostituibile. Il NMAH è in sostanza l'indirizzo web convenzionale dell'host che risolverà, in maniera tradizionale, l'identificatore. Il vero e proprio identificatore, unico, immutabile, globale, è quello che segue l'etichetta «ark:». L'identificatore comprende il *Name Assigning Authority Number* (NAAN), cioè un numero che identifica l'organizzazione che ha assegnato originariamente il nome all'oggetto, seguito a sua volta dal nome assegnato all'oggetto e da un eventuale qualificatore del nome. Vedere l'immagine 7.



7 – Un ARK è un comune URL strutturato in maniera particolare.

La particolarità dell'ARK sta nel connettere tre cose: l'oggetto identificato, i suoi metadati e l'impegno del gestore dell'identificatore circa la sua persistenza. Quando un utente inserisce l'ARK nel campo degli indirizzi del browser, l'identificatore conduce l'utente all'oggetto identificato. Lo stesso ARK, seguito da un singolo punto interrogativo (?) restituisce un breve record di metadati che è leggibile tanto dall'uomo che dalla macchina. Quando l'ARK è seguito da due punti interrogativi (??) i metadati in risposta contengono l'assunzione di responsabilità del gestore.

Questo ci conduce alla enunciazione dei tre requisiti che deve possedere un ARK. Il primo requisito risiede proprio nel legame che l'identificatore stabilisce tra un oggetto e un'assunzione di responsabilità per la manutenzione di quell'oggetto. Il secondo requisito è che l'ARK deve fornire all'utente un link dall'oggetto a una descrizione di questo. A questo scopo l'ARK si avvale del set di metadati METS (*Metadata Encoding and Transmission Standard*) mantenuto dalla Library of Congress.²⁶ Il terzo requisito è quello di garantire l'accesso permanente agli oggetti identificati.

L'ARK trova la sua principale limitazione nel fatto che questo accesso è basato esclusivamente sulla risoluzione tramite il tradizionale sistema DNS. In altri termini:

²⁶ Cfr. <http://www.loc.gov/standards/mets/>.

«Il fatto di premettere «http://» e un NMAH a un ARK, è un modo per creare un identificatore azionabile tramite un metodo che è esso stesso temporaneo. Ipotizzando che un giorno non sia più disponibile l'infrastruttura che supporta il recupero dell'informazione tramite HTTP, l'ARK dovrà essere allora convertito in qualche nuovo tipo di identificatore azionabile».²⁷

Un name-system alternativo al DNS: lo Handle System

Ma la strada più promettente per risolvere il problema dell'identificazione persistente degli oggetti digitali è, probabilmente, quella che consiste nel creare sistemi di risoluzione affatto nuovi, basati su sistemi diversi dal DNS. Il principale esponente di questo approccio è rappresentato dall'*Handle System*.²⁸

Questo sistema è stato inizialmente concepito e sviluppato presso la *Corporation for National Research Initiatives* (CNRI) come parte di una più vasta architettura globale per gli oggetti digitali. La prima implementazione pubblica è stata realizzata presso il CNRI nell'autunno del 1994, grazie a D. Ely. La complessiva architettura globale di oggetti digitali, incluso lo Handle System, è stata poi descritta da R.Kahn e R.Wilensky nel 1995.²⁹ Lo sviluppo è poi proseguito presso il CNRI.³⁰

L'Handle System è un servizio che utilizza un suo specifico protocollo per creare, aggiornare, mantenere e rendere accessibile in modo sicuro un database distribuito per l'associazione permanente tra nomi e oggetti digitali. È stato disegnato per rendere possibili servizi di trasmissione sicura di informazione e condivisione di risorse su network quali appunto Internet. Le applicazioni dello Handle System possono includere servizi di metadati per le pubblicazioni digitali, servizi di gestione di identità per le identità virtuali e/o ogni altra applicazione che richieda la risoluzione e/o l'amministrazione di identificatori unici globali. Naturalmente l'Handle System, come tutte le realizzazioni concrete, è l'espressione di un compromesso tra varie diverse esigenze tecniche e pratiche.

Come configurazione globale del servizio è stato adottato un modello gerarchico. Il top-level del sistema è rappresentato da un singolo servizio handle, meglio conosciuto come *Global Handle Registry* (GHR). I livelli inferiori sono rappresentati da tutti gli altri servizi handle, meglio noti come *Local Handle Services* (LHS). Il GHR

²⁷ John A. Kunze, *The ARK Persistent Identifier Scheme, Internet-Draft*, 19 august 2005, p. 9, <http://www.cdlib.org/inside/dglib/ark/arkspec.pdf>.

²⁸ Sam Sun – Larry Lannom – Brian Boesch, *Handle System Overview*, RFC 3650, novembre 2003, <http://www.ietf.org/rfc/rfc3650.txt>.

Sam Sun – Sean Reilly – Larry Lannom, *Handle System Namespace and Service Definition*, RFC 3651, novembre 2003, <http://www.ietf.org/rfc/rfc3651.txt>.

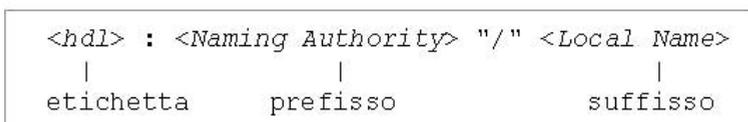
Sam Sun – Sean Reilly – Larry Lannom – Jason Petrone, *Handle System Protocol (ver 2.1) Specification*, RFC 3652, novembre 2003, <http://www.ietf.org/rfc/rfc3652.txt>.

²⁹ Robert Kahn – Robert Wilensky, *A framework for Distributed Digital Object Services*, «D-Lib magazine», 1995, <http://www.cnri.reston.va.us/k-w.html>; anche [hdl:cnri.dlib/tn95-01](http://hdl.cnri.dlib/tn95-01).

³⁰ Cfr. <http://www.handle.net/>.

è il gestore unico di tutte le *naming authorities* appartenenti al sistema. I singoli LHS operano al servizio di una data *naming authority*.

Ogni handle consiste di due parti: la sua *naming authority*, altrimenti nota come prefisso, e un nome locale univoco sotto la *naming authority*, meglio noto come suffisso. Tutto preceduto da un'etichetta che, nel caso di un handle tradizionale, è «hdl:» (immagine 8).



8 – Struttura di un handle.

La *naming authority* identifica l'unità amministrativa responsabile per la creazione e l'amministrazione dell'handle associato. Le *naming authority* sotto lo Handle System hanno struttura gerarchica ma, a differenza del DNS, le *handle naming authorities*, nello schema 8, sono costruite da sinistra verso destra.

Operativamente il sistema funziona così: un utente che debba risolvere un indirizzo handle, interroga con il suo client il GHR; questi invia in risposta l'indirizzo dell'LHS responsabile per la risoluzione di quel dato handle; il client interroga allora LHS in questione; questi a sua volta invia in risposta l'indirizzo della risorsa digitale associata all'handle.

Un aspetto importante dell'Handle System è la sua architettura distribuita. Lo Handle System, nel suo insieme, consiste di vari servizi handle individuali. Ognuno di questi servizi può consistere di uno o più siti di servizio. Ogni sito di servizio, a sua volta, può consistere di uno o più server handle. Tutte le richieste handle, indirizzate a un dato sito di servizio, possono essere ridistribuite, eventualmente, tra i vari server handle. Non ci sono limiti al numero di servizi handle o al numero di siti che possono essere allestiti per ogni servizio o al numero di server per ogni sito.

La comunicazione tra le varie componenti del sistema avviene mediante uno specifico protocollo (lo *Handle System Protocol*), più efficiente sotto molti aspetti rispetto al tradizionale protocollo DNS. In primo luogo le applicazioni DNS non reggono bene le situazioni in cui grandi quantità di dati sono associate con un particolare nome DNS. Inoltre i nomi DNS sono gestiti dall'amministratore di zona di uno specifico livello della gerarchia DNS. Non sono previste modalità per implementare una struttura amministrativa finalizzata specificatamente alla gestione dei nomi. Solo l'amministratore del network può creare o maneggiare i nomi DNS. Ogni handle invece ha il proprio amministratore distinto dall'amministratore del server. Ancora, il protocollo dello Handle System include delle opzioni di sicurezza che garantiscono la riservatezza e l'integrità dei dati durante la trasmissione. Nel DNS questo non c'è.

Riassumendo: «Lo Handle System fornisce un *name service* globale a scopi generali che permette di risolvere e amministrare in maniera sicura gli handle sulla rete pubblica Internet». ³¹ Come è noto, dal punto di vista di un utente finale, per risolvere un URL basta scriverlo dentro lo spazio degli indirizzi del browser e quindi spingere il tasto invio. Anche gli handle si risolvono nello stesso modo. Ma prima occorre adattare il proprio browser Microsoft Internet Explorer o Netscape, scaricando dal sito del CNRI un apposito programmino. Questo si chiama *Handle Resolver* e va installato, con una procedura estremamente semplice, sul proprio PC. Un browser così adattato può risolvere gli handle direttamente. ³²

Non disponendo di un browser adattato, si possono comunque risolvere gli handle utilizzando il normale protocollo HTTP e un server proxy. Il CNRI, ad esempio, rende disponibile un server proxy all'indirizzo <http://hdl.handle.net/>. Risolvere l'handle utilizzando un server proxy consiste nel costruire una normale URL costituita dall'indirizzo del server proxy seguito dall'handle da risolvere.

Nell'immagine 9 possiamo vedere i tre casi riuniti: 1. il normale URL; 2. l'handle da risolvere direttamente; 3. l'handle da risolvere tramite un proxy-server. Tutte e tre le diverse modalità di accesso puntano allo stesso documento in rete, quello già citato di R.Kahn e R.Wilensky.

1.	http://www.cnri.reston.va.us/k-w.html
2.	<code>hdl:cnri.dlib/tn95-01</code>
3.	http://hdl.handle.net/cnri.dlib/tn95-01

9 – Si può accedere ad un dato documento tramite: 1) il tradizionale URL; 2) risolvendo direttamente l'handle; 3) risolvendo l'handle tramite un proxy-server. In grassetto l'handle.

La differenza tra le diverse modalità d'accesso risiede nella persistenza: se dovesse cambiare la localizzazione del documento, l'handle, a condizione che le informazioni nell'LHS vengano aggiornate, continuerà a operare. L'URL no. L'unica limitazione per l'handle sta nell'indirizzo del server proxy che, essendo un normale indirizzo DNS, può variare anch'esso. In tal caso anche gli indirizzi come quello del punto 3. non saranno più validi. Ma l'handle al punto 2. continuerà comunque a funzionare correttamente.

Il Digital Object Identifier (DOI)

Attualmente l'implementazione più importante dell'Handle System è rappresentata dal *Digital Object Identifier* (DOI). Il DOI System è gestito dall'*International Doi Foundation* (IDF)³³, un consorzio aperto che comprende partner sia commerciali

³¹ *Handle System Namespace and Service Definition*, RFC 3651 cit., p. 1.

³² Lo *Handle Resolver* si scarica da http://www.handle.net/resolver/win_95_98_NT.html.

³³ <http://www.doi.org>.

che non-commerciali. All'IDF fanno capo svariate agenzie regionali di registrazione DOI che finora hanno assegnato alcuni milioni di DOI negli Stati Uniti, in Europa, in estremo oriente.

Il DOI è a tutti gli effetti un handle e l'IDF una *handle naming authority*. Un DOI si riconosce per la prima parte del prefisso che è sempre 10; ad esempio: 10.1045/xxxx è un DOI. La seconda parte del prefisso (1045) è l'identificativo della specifica agenzia di registrazione responsabile di quel particolare DOI; il suffisso è il nome della risorsa identificata in maniera permanente da quel DOI (xxxx). Il nome può essere rappresentato da qualsiasi sequenza di caratteri. Si possono usare per i nomi anche sistemi di codifica già esistenti. Ad esempio 10.1000/ISBN1-900512-44-0 è un DOI valido che incorpora un codice ISBN. Recentemente il DOI è diventato uno standard NISO (*National Information Standards Organization*).³⁴

Anche un DOI può essere risolto direttamente utilizzando lo stesso Handle Resolver indicato prima, oppure via proxy-server, utilizzando il sistema allestito dall'IDF; a questo scopo si usa l'URL <http://dx.doi.org/>.³⁵ Nell'immagine 10 vediamo un esempio, analogo al n. 8, che espone tre diverse modalità di accesso a un dato documento (nella fattispecie all'editoriale di D-Lib Magazine di luglio/agosto 2005): 1. tramite l'URL convenzionale; 2. risolvendo direttamente il DOI dell'editoriale; 3. risolvendo il DOI tramite il proxy-server della International Doi Foundation.

1.	http://www.dlib.org/dlib/july05/07editorial.html
2.	doi:10.1045/july2005-editorial
3.	http://dx.doi.org/10.1045/july2005-editorial

10 – Si può accedere a un documento tramite: 1) l'URL convenzionale; 2) risolvendo direttamente il DOI; 3) risolvendo il DOI tramite il proxy-server della IDF. In grassetto il DOI.

Una caratteristica importante del DOI è quella di essere un sistema a pagamento. La spiegazione che viene data di questa circostanza è che «come per la registrazione dei nomi di dominio, anche l'assegnazione del DOI richiede un pagamento e l'impegno a seguire standard e regole definite. Questo non rende il sistema chiuso, o commerciale, ma lo rende gestibile. La DOI Foundation è un'organizzazione non-profit, non un operatore commerciale; tuttavia il sistema ha dei costi che vanno coperti. La persistenza è una funzione delle organizzazioni, non una tecnologia: per supportare un sistema di identificazione persistente, occorre che esista una organizzazione persistente».³⁶

³⁴ NISO Z39.84-200X.

³⁵ Cfr. http://www.doi.org/doi_proxy/index.html.

³⁶ *The DOI Handbook* cit., p. 24-25.

Per registrare un DOI occorre rivolgersi a una apposita agenzia. La più importante in assoluto, con oltre 17.944.074 DOI assegnati, è certamente CrossRef.³⁷ Questa agenzia consente di assegnare i DOI ai seguenti materiali:

- periodici/documenti di lavoro: titolo del periodico, volume, numero, articolo;
- libri: collana, titolo, capitolo, voce;
- atti di convegni: titoli multi-volume, titoli, elaborati;
- componenti: sub-voci di articoli di periodici; voci/capitoli di libri e di atti di convegni comprese le figure, le tavole e gli aggiornamenti.

CrossRef si sta attualmente attrezzando per includere tra i materiali ai quali può essere assegnato un DOI anche le tesi, le dissertazioni e i record di database. CrossRef è un'associazione indipendente di editori. La piattaforma di questa organizzazione funziona come un network di citazioni incrociate che copre milioni di articoli e altri elementi di contenuto appartenenti a varie centinaia di editori accademici e professionali. Recentemente CrossRef ha integrato la sua piattaforma con un resolver OpenURL liberamente disponibile al pubblico, enfatizzando così le possibilità di navigazione all'interno del proprio materiale registrato.³⁸

L'agenzia di registrazione DOI ufficiale per l'Europa è mEDRA - multilingual European DOI Registration Agency - nominata ufficialmente dall'IDF il 1° luglio 2003. L'agenzia opera attraverso una società, mEDRA srl, costituita dall'Associazione Italiana Editori insieme al Cineca.³⁹ L'adesione al servizio richiede il pagamento di una quota annua.⁴⁰ Tale quota include la registrazione di un numero variabile di DOI, a seconda della fascia di pagamento. In caso di eccedenza dei DOI previsti dalla quota annua, è previsto un costo unitario per i DOI eccedenti, anch'esso rapportato alla fascia della quota annua. La persistenza dei DOI registrati è garantita per almeno 5 anni dall'ultimo anno di adesione. Vedere la tabella 11.

Fascia	Quota fissa	DOI inclusi	Costo unitario DOI extra
1	400	100	2,00
2	600	200	1,50
3	900	400	1,20
4	1.350	750	1,00
5	2.000	1.500	0,80
6	3.000	3.000	0,70
7	4.500	6.000	0,60

11 – Fasce di pagamento DOI secondo l'agenzia per l'area europea mEDRA (in euro).

³⁷ Cfr. <http://www.crossref.org>.

³⁸ OpenURL è un tipo di URL che contiene metadati di uso bibliografico. OpenURL è un sistema originariamente ideato da Herbert van de Sompel, un bibliotecario dell'Università di Ghent. Cfr. <http://www.crossref.org/01company/pr/press081505.htm>.

³⁹ Cfr. <http://www.medra.org/it/index.htm>.

⁴⁰ Cfr. http://www.medra.org/it/medra_appl.htm.

Naturalmente, quando un DOI viene assegnato a un dato oggetto, nel database della agenzia di assegnazione deve essere inserita una descrizione di quell'oggetto. Il DOI Handbook, a questo riguardo, prescrive un set minimo di metadati denominato *DOI Kernel Metadata Declaration*. Ogni risorsa identificata con il DOI deve essere descritta secondo le specifiche di questo insieme minimo di metadati.⁴¹ Ma oltre al *kernel*, le agenzie di registrazione possono implementare ulteriori schemi di metadati, soprattutto a fini di interoperabilità tra piattaforme e servizi differenti. mEDRA, a questo riguardo, si avvale di un set di metadati, denominato ONIX (*Online Information Exchange*), che è uno standard internazionale per rappresentare le informazioni commerciali su libri, riviste e video in formato elettronico. Questo set è mantenuto da EDItEUR⁴², un gruppo internazionale che coordina lo sviluppo e la promozione degli standard per il commercio elettronico nell'ambito dell'editoria libraria e dei periodici.

mEDRA ha selezionato, all'interno dell'intera specifica di metadati ONIX, otto differenti sottoinsiemi di metadati per scopi specifici. Infine va osservato che anche mEDRA gestisce un proxy-server per risolvere i DOI (<http://dx.medra.org/>).

L'identificazione persistente è un problema di Internet Governance

In conclusione possiamo dire che la soluzione tecnica più promettente per il problema dell'identificazione persistente degli oggetti digitali sembra viaggiare al di fuori dei binari costituiti dal DNS e dal protocollo HTTP. Ma il mondo degli sviluppatori di Internet è riluttante ad abbandonare questa strada. Una segno di ciò può essere individuato nel fatto che, sebbene lo Handle System sia stato riconosciuto come standard Internet, lo stesso non è ancora avvenuto per il DOI. L'apparente paradossale conservatorismo che, per molti versi, caratterizza la cerchia degli sviluppatori di Internet, è ben espresso dal cosiddetto *principio di robustezza*, una massima formulata nel 1981 da J. Postel, uno degli sviluppatori storici di Internet: «Le implementazioni del TCP seguiranno un principio generale di robustezza: siate conservativi in ciò che fate, siate liberali in ciò che accettate da altri».⁴³

Forse però, come ha osservato N. Paskin, dietro il conservatorismo in materia tecnologica, possono nascondersi anche interessi di vario genere: «L'attuale prevalenza nell'*Internet Governance* e, cosa forse ancor più importante, nel finanziamento di Internet, di organizzazioni che fanno affidamento su di un unico sistema di denominazione, il sistema a nomi di dominio (un meccanismo che rende particolarmente difficile identificare i contenuti digitali indipendentemente dalla localizza-

⁴¹ *The DOI Handbook* cit., Appendice 6, p. 143.

⁴² Cfr. <http://www.editeur.org>.

⁴³ Jonathan Bruce Postel, *Transmission Control Protocol. Darpa Internet Program. Protocol Specification*. RFC 793, Sept. 1981, pag. 13, <http://www.ietf.org/rfc/rfc793.txt>.
J. Postel (6.8.1943 – 16.10.1998) ha contribuito alla redazione di oltre 200 RFC.

zione) può essere all'origine delle difficoltà nell'introdurre meccanismi di denominazione alternativi e complementari".⁴⁴ Ma, d'altra parte, è improbabile che il DNS possa durare in eterno. Infatti, come ha sottolineato K. N. Cukier, corrispondente tecnologico dell'*Economist*, in un convegno sull'*Internet Governance* nel maggio di quest'anno: «Il sistema di nomi di dominio, numeri IP, *root server* e identificatori di protocollo non è una tecnologia statica, bensì una capace di evolvere in una forma migliore. Di conseguenza il sistema corrente non deve essere considerato sacrosanto ma suscettibile di innovazioni".⁴⁵

A questo punto, quindi, il problema dell'identificazione persistente degli oggetti digitali sfuma nella tematica più generale dell'*Internet Governance*, intendendo con questo termine, l'insieme delle strutture e delle organizzazioni che, con le loro strategie politiche e le loro reciproche interazioni, di fatto determinano, oggi, l'effettivo concreto sviluppo di Internet.

⁴⁴ Norman Paskin, *Two practical examples of issues in internet governance*, Internet Governance Forum, Oxford Internet Institute, May 6th 2005, <http://www.doi.org/topics/050504011DiscussionForum.pdf>.

⁴⁵ Kenneth Neil Cukier, *Slouching Towards Geneva: Ten appreciated Axioms of Internet Governance*, Internet Governance Forum, Oxford Internet Institute, May 6th 2005.

RIFERIMENTI BIBLIOGRAFICI

- Tim Berners-Lee – Rot T. Fielding – Larry Masinter. *Uniform Resource Identifier (URI): Generic Syntax*. RFC3986, gennaio 2005, <http://www.ietf.org/rfc/rfc3986.txt>.
- Tim Berners-Lee – Larry Masinter – Mark McCahill (Eds.). *Uniform Resource Locators (URL)*. RFC 1738, dicembre 1994, <http://www.ietf.org/rfc/rfc1738.txt>.
- Kenneth Neil Cukier. *Slouching Towards Geneva: Ten appreciated Axioms of Internet Governance*, Internet Governance Forum, Oxford Internet Institute, 6 maggio 2005, <http://www.cukier.com/writings/cukier-oi-netgov-may05.pdf>.
- Robert E. Kahn – Vinton G. Cerf. *What Is The Internet (And What Makes It Work)*, dicembre 1999, http://www.cnri.reston.va.us/what_is_internet.html
- Robert E. Kahn – Robert Wilensky. *A framework for Distributed Digital Object Services*. «D-Lib magazine», 1995.
- John A.Kunze. *Towards Electronic Persistence Using ARK Identifiers*, University of California, luglio 2005, <http://www.cdlib.org/inside/diglib/ark/arkcdl.pdf>.
- Michael Mealling – Ray Denenberg (Eds.). *Report from the Joint W3C/IETF URI Planning Interest Group: Uniform resource Identifiers (URIs), URLs., and Uniform Resource Names (URNs): Clarifications and Recommendations*. RFC 3305, agosto 2002, <http://www.ietf.org/rfc/rfc3305.txt>.
- Paul Mockapetris. *Domain names - Implementation and Specification*. RFC 883, novembre 1983. <http://www.ietf.org/rfc/rfc883.txt>.
- Norman Paskin. *Digital Object Identifiers for scientific data*. «Data Science Journal», Volume 4, 18 marzo 2005.
- Norman Paskin. *DOI. A 2003 Progress report*. «D-Lib Magazine», giugno 2003, vol.9, num.6.
- Norman Paskin. *On Making and Identifying a "Copy"*. «D-Lib Magazine», gennaio 2003, vol.9, num.1.
- Norman Paskin (Ed.). *The DOI Handbook. Edition 4.2.0*, International Doi Foundation, 2005, <http://dx.doi.org/10.1000/186>.
- Norman Paskin. *Two practical examples of issues in internet governance*, Internet Governance Forum, Oxford Internet Institute, 6 maggio 2005, <http://www.doi.org/topics/0505040IIDiscussionForum.pdf>.
- Jonathan Bruce Postel. *Transmission Control Protocol. Darpa Internet Program. Protocol Specification*, RFC 793, settembre 1981, <http://www.ietf.org/rfc/rfc793.txt>.
- Sam Sun – Larry Lannom – Brian Boesch. *Handle System Overview*. RFC 3650, novembre 2003, <http://www.ietf.org/rfc/rfc3650.txt>.
- Sam Sun – Sean Reilly – Larry Lannom. *Handle System Namespace and Service Definition*. RFC 3651, novembre 2003, <http://www.ietf.org/rfc/rfc3651.txt>.
- Sam Sun – Sean Reilly – Larry Lannom – Jason Petrone. *Handle System Protocol (ver 2.1) Specification*. RFC 3652, novembre 2003, <http://www.ietf.org/rfc/rfc3652.txt>. <http://www.cnri.reston.va.us/k-w.html>; anche hdl.cnri.cdlib.org/tn95-01.
- Elaine Svenonius (a cura di). *The conceptual foundations of descriptive cataloging*, San Diego: Academic Press, 1989.
- Elaine Svenonius. *The intellectual foundation of information organization*, Cambridge (Mass.): MIT Press, 2000.
- Zeno Tajoi. *Doi: uno strumento per costruire la biblioteca digitale*. «Bollettino AIB», 45, n.1, marzo 2005.
- Giuseppe Vitiello. *Identifiers and Identification Systems*. «D-Lib Magazine», gennaio 2004, vol.10, num.1.
- Patrick Wilson. *The Second Objective. The conceptual foundations of descriptive cataloging*, San Diego: Academic Press, 1989.