

# Citazione delle pubblicazioni digitali

Anna Maria Tammaro

Università degli studi di Parma  
annamaria.tammaro@unipr.it

*Opinioni a confronto e problemi aperti*

## Introduzione

Quando parliamo di comunicazione scientifica intendiamo una catena di autori che, superando i limiti di tempo e di spazio, collaborano al continuo sviluppo della conoscenza. Ogni ricerca inizia dalla letteratura sull'argomento, per evitare duplicazioni di ricerche già fatte ed anche per valutare la qualità e il rigore scientifico di quanto viene affermato. Un'esigenza essenziale della comunicazione scientifica è quindi la possibilità di verificare quello che l'autore ha citato nel suo lavoro. Questa esigenza rimane essenziale nell'ambito digitale, ma la citazione subisce i cambiamenti che l'evoluzione (o la rivoluzione) della comunicazione scientifica sta vivendo nel passaggio al digitale.<sup>1</sup>

La citazione è un sistema che rende possibile l'identificazione e l'accesso alla risorsa citata. Per la costruzione di questo sistema, occorre trovare sia un modello di descrizione del documento sia un modello di organizzazione delle risorse citate. Il sistema delle citazioni è sufficientemente funzionale nel mondo delle pubblicazioni a stampa, ma non è ugualmente efficace nel mondo digitale.

La comunicazione scientifica sempre di più si basa su risorse digitali che sono presenti in Internet e che vengono citate con la tipica modalità del web, cioè l'URL della pagina web. Il problema tuttavia è che spesso le risorse digitali nel

web non sono persistenti ed è diffusa la situazione in cui, anche dopo poco tempo dall'ultimo accesso, l'URL citata non è più attiva (il comune "errore 404"). Un altro problema assai frequente è che il contenuto della risorsa digitale citata spesso cambia nel tempo. Come possiamo, quindi, adattare le regole di citazione all'ambito digitale? Come possiamo inoltre trovare un'organizzazione efficace delle risorse digitali che sono citate?

## I sistemi di citazione

Esistono diverse opinioni riguardo ai sistemi di citazione in ambito digitale e quindi possiamo evidenziare stili di citazione e approcci al problema contrapposti, da diversi punti di vista.

Un primo approccio, che è portato avanti soprattutto dalla comunità dei bibliotecari, cerca di adattare le tradizionali norme di citazione bibliografica all'ambito digitale (Gnoli 2000; Revelli 2002; Ridi 1995). Con la nascita dei documenti digitali sono state infatti realizzate delle aggiunte alle regole di citazione normalmente usate per le pubblicazioni a stampa. Possiamo ad esempio indicare lo standard *ISO 690-2 Information and documentation – Bibliographic references – Part 2: Electronic documents or arts thereof*.<sup>2</sup> Altri standard molto diffusi per la citazione delle pubblicazioni a stampa, come le regole di citazione dell'APA o di MLA,

sono state adattate ai documenti elettronici.<sup>3</sup> In particolare, questo primo approccio al problema della citazione in ambito digitale identifica il documento digitale con un documento che ha le stesse caratteristiche di un documento tradizionale, però su supporto digitale. In questo caso, ci si riferisce a documenti che quasi sempre hanno una copia (o un originale) su supporto cartaceo, che segue la tradizionale organizzazione. Questo approccio considera importante educare gli utenti per migliorare le loro capacità di saper identificare e recuperare i documenti, in particolare gli articoli scientifici, i libri di testo e il materiale didattico, e altre risorse digitali citabili. Un altro approccio, invece, considera importante la differenza dei documenti digitali rispetto alle pubblicazioni a stampa, per affermare che le regole di citazione tradizionali non sono più adeguate e vanno trovate regole diverse. Ad esempio: com'è possibile identificare il documento digitale nella sua granularità e navigabilità? Come va citato un documento ipertestuale? Di fronte alle nuove tipologie di documenti digitali presenti nel web, si cerca di risolvere i problemi della citazione usando le stesse tecnologie digitali, e la citazione diventa semplicemente un collegamento (o link): la caratteristica di questo approccio è quella di usare la localizzazione della risorsa digitale come sistema di citazione. Tuttavia bisogna avvertire

che questo sistema, che tende a semplificare il sistema di citazione, non è stabile e quindi non è affidabile. Per rendere possibile il collegamento a una risorsa, questa deve avere un identificatore univoco e persistente. Gli aspetti tecnici della citazione hanno trovato diverse soluzioni, come lo standard ISO Open URL, l'identificatore PURL plus, DOI, che definisce anche i diritti di proprietà associati alla risorsa, e nel contesto di Web 02, infine, possono essere considerati anche i *social tagging* (Berners Lee 2004). È interessante notare che PURL plus definisce gli elementi minimi necessari per la citazione: il dominio, definito come l'istituzione o il contesto dell'autore, l'autore o creatore della risorsa, il contenuto, la data, la misura del file. In futuro, alcuni sistemi di ricerca della rete, come ad esempio i motori di ricerca, potranno funzionare meglio dello stesso identificatore URI<sup>4</sup> (Jasco 2005; Lawrence and Giles 1998). Il problema della citazione in ambito digitale è quindi ancora senza soluzione, se non una pragmatica. Quando si citano documenti digitali che hanno una corrispondente versione a stampa, come ad esempio i periodici elettronici e anche alcuni pre-print negli archivi *open access*, può essere applicato il sistema tradizionale di citazione, adattato ai documenti digitali. Quando sono citati altri documenti digitali che esistono solo in versione web e hanno una forma dinamica, come ad esempio le pagine web, per questi sono necessari nuovi sistemi di citazione. Tuttavia come può essere risolta l'esigenza di accesso a tutti i documenti citati che hanno rilevanza per la comprensione di un lavoro scientifico? Per arrivare a un nuovo modello di citazione delle pubblicazioni scientifiche, si dovrà partire da una migliore definizione di documento digitale.

### *Cos'è un documento digitale?*

Molto prima dell'era digitale, il concetto di documento è stato a lungo dibattuto e convivono diversi approcci. I documenti sono stati considerati, alternativamente o in combinazione, come semplici canali di comunicazione, o come supporti di memorizzazione, o recentemente come segni e significati da comunicare. Pedauque (Pedauque 2003) sintetizza gli approcci al concetto di documento in una tassonomia, basata su tre aspetti: Forma, Segno e Supporto. Per l'autore, i ricercatori interessati alla Forma dei documenti hanno interesse alla struttura di questi, a prescindere dal supporto di memorizzazione. La ricerca sui documenti che si concentra sui Segni ha interesse al significato e alla rappresentazione del contenuto, come ad esempio con l'indicizzazione dei soggetti. La ricerca sui documenti più interessata al Supporto si concentra sul ruolo sociale dei documenti, come traccia di una comunicazione che esiste al di là del tempo e dello spazio. I tre aspetti possono anche essere combinati, con la prevalenza di uno o dell'altro. Il documento, comunque inteso, richiede una ri-definizione in ambito digitale, in quanto ha perso non solo la sua materialità ma pone diverse domande circa la sua granularità, la garanzia di autenticità e la sua stabilità nel tempo. Fino all'avvento del digitale, i documenti considerati nella citazione degli articoli scientifici sono stati soprattutto libri o testi scritti; attualmente esistono documenti non tradizionali, come:

- *il documento multimediale*: il documento digitale rappresenta la convergenza di ogni tipo di media. Questo consente all'utente di poter scegliere tra diverse tipologie di rappresentazione dell'informazione contenute nello stesso documento digitale;

– *il documento ipertestuale*: è sicuramente la tipologia digitale più importante dal punto di vista dell'utente. Rappresenta la possibilità del documento digitale di essere collegato ad altri documenti, sia dall'autore che dal lettore. Il documento digitale non è quindi più un documento statico e compiuto, come quello tradizionale, ma consente un'interazione e possibilità pressoché infinite di collaborazione;

– *la visualizzazione a tre dimensioni*: sono modelli interattivi con cui l'utente può interagire in modo avanzato;

– *i documenti interattivi*: alcune banche dati e periodici elettronici che consentono ad esempio delle annotazioni o altre funzionalità di partecipazione da parte degli utenti. Il documento digitale può quindi essere dinamico e anche collaborativo, a differenza del documento tradizionale.

Con un approccio pragmatico, sono state tentate varie definizioni che riguardano il diverso ruolo dei documenti dal punto di vista degli utenti.

È importante ad esempio definire la granularità del documento e anche conoscere i diversi detentori di proprietà intellettuale dei singoli elementi che lo compongono (Kircz 1998). Nel mondo a stampa, il codice ISBN era sufficiente per identificare il libro o il codice SICI per identificare l'articolo. Nel mondo digitale occorre identificare il singolo capitolo dell'e-book o la foto contenuta in un articolo, anche per consentire la navigazione diretta a questa singola parte, se consentita dalle licenze di accesso. In un approccio solo tecnologico, la definizione standard di documento digitale che dà SGML è la seguente:

A collection of information that is processed as a unit.

Per gli scopi di SGML, che si con-

centra sulla struttura del documento, ci si riferisce a un'organizzazione gerarchica e logica – non fisica – del documento. Quello che interessa evidenziare sono gli elementi costitutivi e le relazioni che caratterizzano il tipo di documento, o DTD (Document Type Definition). Per semplificare: un documento potrebbe quindi pragmaticamente dirsi un file, o una sequenza di *bits*. Da un punto di vista tecnico, quello che è necessario è una struttura standard del documento, utile per essere interoperabile in un'architettura di rete basata sui servizi (come ad esempio SOAP e RTF), invece di un'architettura gerarchica e spesso chiusa, come quella adottata dai sistemi di automazione della biblioteca (Library Management Systems).

In un'interpretazione estesa di documento digitale, coniugata all'ipertesto e all'interattività, potremmo arrivare a indicare il "docuverso", termine coniato da Nelson per l'insieme dei documenti digitali esistenti in rete, come tipologia esemplare di documento attualmente non citabile. Da un punto di vista bibliotecario, potremmo infatti distinguere i documenti citabili, come quelli che fanno parte di collezioni digitali e di biblioteche digitali, e documenti non citabili, quelli che invece non lo sono. Come eventuali eccezioni a questa distinzione, si potrebbero includere nel primo gruppo le risorse digitali controllate dagli editori, e accessibili sulla base di accordi specifici. Occorre tuttavia essere certi che la preservazione della risorsa digitale sia prevista dalle licenze di uso. Il concetto di documento citabile, che può includere sia documenti statici che dinamici, è quindi legato soprattutto alla gestione dei documenti digitali in depositi garantiti (*trusted repositories*) che ne assicurano la preservazione. Bisogna evidenziare tuttavia che, insieme alla crescita dei documenti

digitali disponibili, è nato il fenomeno collegato di diversi vincoli di accesso, posti dai detentori dei diritti di proprietà intellettuale.

### *Autore e lettore*

Per risolvere il problema del sistema di citazione dovremmo chiederci prioritariamente: a chi serve la citazione?

La citazione è innanzitutto necessaria per gli autori che vogliono comunicare ai lettori: "Io ho usato questa risorsa digitale". Gli autori pubblicano essenzialmente perché vogliono il riconoscimento del loro lavoro, e così acquistare autorevolezza e fare carriera. Non è un caso che il sistema della citazione sia stato correlato da sempre alla qualità, che viene riconosciuta alla pubblicazione (Garfield 1979). Fin dal primo periodico scientifico, il "Journal des Savants", questo si presentava come un sistema di selezione critica, con lo scopo di registrare e validare il valore scientifico di una particolare ricerca pubblicata. Il fattore d'impatto, e altri strumenti bibliometrici come le leggi di Bradford e di Lotka, sono divenuti gli standard per la valutazione della ricerca scientifica.

Tuttavia anche il concetto di autore sta cambiando (Hitchcock 1997). Le citazioni servono soprattutto ai lettori: questi vogliono l'assicurazione di poter ritrovare i documenti citati navigando in rete attraverso i collegamenti. Inoltre gli utenti si aspettano la certificazione delle pubblicazioni citate. La prima funzione che viene richiesta dai lettori riguarda il bisogno di avere a disposizione la letteratura scientifica. La seconda funzione va incontro alla necessità, che nasce al crescere della quantità e della disponibilità di diverse versioni delle pubblicazioni digitali, di certificare la versione e la non manipolazione delle pubblicazioni digitali. Si tratta di questioni correlate,

poiché chi fa una citazione indica, al contempo, una determinata versione. Al cuore del problema delle citazioni sta quindi la nozione di qualità del documento digitale.

Molti autori sono convinti che le pubblicazioni digitali abbiano un valore minore rispetto a quelle su carta. Al contrario di questa opinione corrente, molti documenti digitali in Internet, come quelli accessibili nei depositi istituzionali, sono organizzati con particolari garanzie di autenticità e qualità. In particolare, il movimento e le iniziative che oggi si mobilitano per l'accesso aperto (Open Access) hanno esplorato nuove soluzioni per disseminare la produzione scientifica, in particolare quella finanziata con fondi pubblici (Di Donato 2004), e ricerche recenti dimostrano un crescente numero di citazioni per i lavori ad accesso aperto (Bauer – Bakkalbasi 2005; Vaughan – Shaw 2005).

### **Organizzazione dei documenti digitali**

Anche limitando il sistema della citazione ai soli documenti digitali che sono gestiti da depositi certificati, quali sono le differenze di organizzazione dei documenti che possono assicurare la loro permanenza e la loro autenticità? Le istituzioni che gestiscono la produzione e/o la gestione di contenuti digitali hanno la responsabilità di un'organizzazione affidabile che sia trasparente agli utenti. Un problema organizzativo è: come rendere noto agli utenti che il documento è stato spostato? Oppure che il contenuto è stato cambiato? Come vengono resi accessibili i documenti? Come vengono gestiti i documenti dinamici? Anche il Web semantico e le ontologie sono importanti per i depositi di contenuti digitali. Tutto questo ha un impatto sulla neutralità dell'or-

ganizzazione della conoscenza, rispetto alle procedure tradizionali. Possiamo indicare solo brevemente alcune delle differenze. È importante una buona organizzazione dei metadati per le risorse digitali, in quanto senza metadati una risorsa digitale non solo è difficilmente identificabile ma non si può nemmeno gestire. Tuttavia bisogna anche affermare che la catalogazione delle risorse digitali costa più di quella delle risorse tradizionali, se non attuata attraverso sistemi semiautomatici. La citazione dei documenti ha delle implicazioni per la conservazione dei supporti di memorizzazione e per l'autenticità della risorsa digitale. La problematica della preservazione e il problema del deposito legale vanno quindi considerati connessi alle regole di citazione. Inoltre va considerata anche la problematica dell'autenticità. Costruire dei falsi, manipolando i file, è ancora più facile nel mondo digitale che nel mondo a stampa. Una particolare debolezza, che può essere considerata per altri versi anche un punto di forza del documento digitale, è proprio il possibile ri-uso e la manipolazione che altri autori possono fare del documento digitale. Possiamo qui evidenziare la necessità di una maggiore convergenza tra il sistema archivistico e quello bibliotecario per l'organizzazione dei documenti digitali, su cui ancora manca sufficiente discussione.

Quali possibili soluzioni al sistema della citazione sono state indicate dal punto di vista tecnologico? Queste comprendono l'harvesting dei dati a richiesta (*on demand*) oppure, se consentito dalle licenze di accesso, l'incapsulamento del documento citato insieme a quello che lo cita (Lawrence, Giles, and Bollacker 1999). Un'altra possibilità è l'autocitazione (Cameron 1997).<sup>5</sup> In questo caso occorrerà avere a disposizione dei depositi

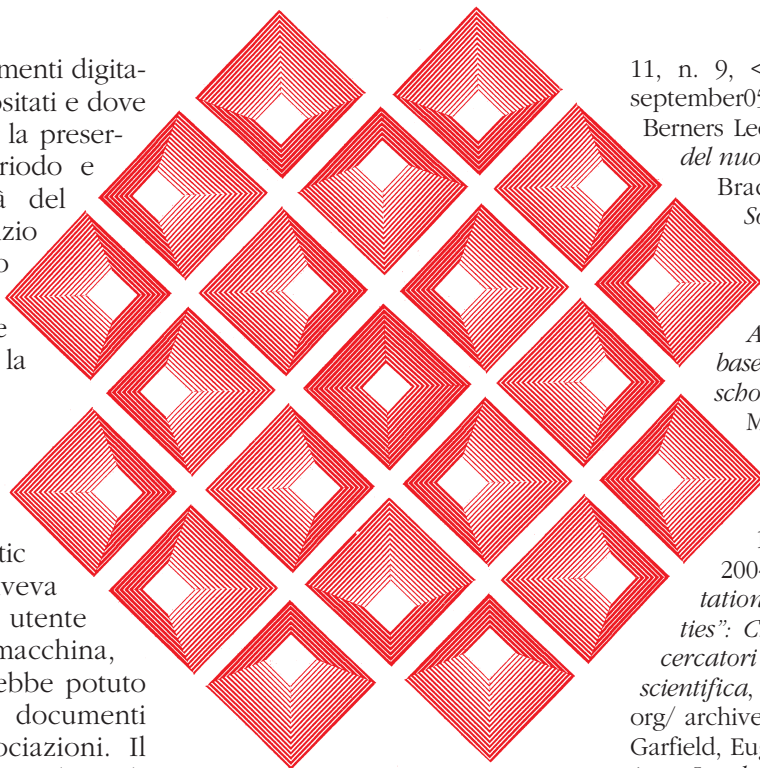
certificati, dove i documenti digitali possano essere depositati e dove venga anche garantita la preservazione di lungo periodo e certificata l'autenticità del documento. Tale servizio potrebbe essere gestito da un'agenzia nazionale o da un'istituzione che possa garantire la preservazione.

### Conclusioni

Nel 1945 su "The Atlantic Monthly", Bush descriveva lo scenario in cui un utente ideale di una teorica macchina, di nome Memex, avrebbe potuto archiviare e reperire documenti collegati tramite associazioni. Il Memex funzionava in modo molto simile a quanto fa oggi il web, e ha la medesima finalità, cioè costruire sistemi informativi più ricchi e intelligenti. Che la creazione di conoscenza si fondi sul concetto di "collegamento" non è una novità; come scrive l'inventore del web:

la comunità dei ricercatori usa da sempre collegamenti del genere tra documenti cartacei: tavole dei contenuti, indici analitici, bibliografie e sezioni di consultazione e rimandi sono tutti link ipertestuali (Berners Lee 2004).

La novità nell'ipotesi di Bush e dei suoi successori è che essi si propongono di *costruire* un sistema informativo basato sulle associazioni, quello che Ted Nelson definirà nel 1981 "Hypertext", una griglia in espansione in grado, potenzialmente, di raccogliere in un unico sistema tutti i testi della letteratura mondiale. Possiamo quindi dire che il sistema delle citazioni è centrale fin dalla stessa visione originaria della biblioteca digitale e del web, ma con l'avvio del conte-



sto digitale ha posto il problema di nuovi documenti e di una nuova organizzazione della conoscenza. La discussione è vivace, ma ancora non è possibile delineare un sistema compiuto e organizzato di citazioni in ambito digitale. In questo periodo di transizione è necessario trovare un equilibrio tra i sistemi tradizionali di citazione e i nuovi sistemi che stanno nascendo. Il cambiamento è appena iniziato ed è necessaria una ricerca più approfondita per trovare delle soluzioni a tutte le problematiche sollevate. A conclusione di questo articolo, nel dubbio se fare una Web-bibliografia distinta dai riferimenti bibliografici, si è alla fine optato per un'unica lista, pur consapevole della mancanza di coerenza dei criteri applicati.

### Riferimenti bibliografici

APA Style. <<http://www.apa.org/journals/webref.html>>.  
Bauer, Kathleen and Bakalbas, Nisa. 2005. *An examination of citation counts in a new scholarly communication environment*, "D-Lib Magazine",

11, n. 9, <<http://www.dlib.org/dlib/september05/bauer/09bauer.html>>.

Berners Lee, Tim 2004. *L'architettura del nuovo web*, Milano, Feltrinelli.

Bradford, Samuel C. 1934.

*Sources of information of specific subjects*, "Engineering", 137, p. 85-86.

Cameron, Robert D. 1997.

*A universal citation database as a catalyst for reform in scholarly communication*, "First Monday", 2, n. 4. <[http://www.firstmonday.dk/issues/issue2\\_4/cameron/](http://www.firstmonday.dk/issues/issue2_4/cameron/)>.

Di Donato, Francesca.

2004. *Verso uno "european citation index for the humanities": Che cosa possono fare i ricercatori per la comunicazione scientifica*, ELIS, <<http://eprints.rclis.org/archive/00002550/>>.

Garfield, Eugene 1979. *Citation indexing: Its theory and application*. In *Science, technology, and humanities*, New York, John Wiley & Sons.

Gnoli, Claudio 2000. *Le citazioni bibliografiche*, AIB-WEB, <<http://www.Aib.it/aib/contr/gnoli2.htm>>.

Hitchcock, S. et al. 1997. *Citation linking: Improving access to online journals*, in *ACM International Conference on Digital Libraries*, 2nd, New York, ACM Press, p. 115-122.

*ISO 690-2 Information and documentation – Bibliographic references – Part 2: Electronic documents or arts thereof*, <<http://www.nlc-bnc.ca/iso/tc46sc9/standard/690-2e.htm>>.

Jasco, Peter 2005. *Scholar and web of science*, "Current Science", 89, n. 9, p. 1537-1547.

Kircz, Joost G. 1998. *Modularity: The next form of scientific information presentation?*, "Journal of documentation", 54, March, p. 210-235.

Lawrence, Steve and Giles, C. Lee 1998. *Searching the world wide web*, "Science", 280, n. 5360, p. 98-100.

Lawrence, Steve, Giles, C. Lee, and Bollacker, Kurt 1999. *Digital libraries and autonomous citation indexing*, "IEEE Computer", 32, n. 6, p. 67-71.

Lotka, Alfred James 1926, *The frequency distribution of scientific productivity*, "Journal of the Washington Academy of Sciences", 16, n. 12, p. 317-323.

Pedauque, R.T. 2003. *Document: Form*,

*sign and medium, as reformulated for electronic documents*, Working paper, version 3, July 8, <[http://archivesic.ccsd.cnrs.fr/documents/archives0/00/00/05/94/sic\\_00000594\\_01/sic\\_00000594.pdf](http://archivesic.ccsd.cnrs.fr/documents/archives0/00/00/05/94/sic_00000594_01/sic_00000594.pdf)>.

Revelli, Carlo 2002. *Citazione bibliografica*, Roma, AIB.

Ridi, Riccardo 1995. *Citare Internet*, "Bollettino AIB", 35, n. 2, p. 211-220.

Vaughan, Liwen and Shaw, Debora 2005. *Web citation data for impact assessment*, "Journal of the American Society for Information Science", 10, n. 56, p. 1075-1087.

(Siti web visitati l'8 aprile 2007.)

### Note

<sup>1</sup> L'articolo prende spunto dalla discussione avvenuta nel dicembre 2006 a Firenze in occasione di "Ex-cite: an unconference on citation in digital environment", un evento organizzato dal Progetto europeo DPE Digital Preservation Europe. Problematiche e tendenze correlate al problema delle citazioni delle pubblicazioni digitali so-

no state sviscerate in modo non convenzionale, usando l'innovativa struttura dell'*unconference* o, potremmo dire, della non-conferenza. Questo modello di discussione si basa sulla legge di Dave Winer: "The sum of the expertise of the people in the audience is greater than the sum of expertise of the people on stage".

<sup>2</sup> <<http://www.nlc-bnc.ca/iso/tc46sc9/standard/690-2e.htm>>.

<sup>3</sup> <[http://english.ttu.edu/kairos/1.2/inbox/mla\\_archive.html](http://english.ttu.edu/kairos/1.2/inbox/mla_archive.html)> e <<http://www.apa.org/journals/webref.html>>.

<sup>4</sup> Il termine URI (Uniform Resource

Identifier) definisce un identificatore che include sia l'URL (Uniform Resource Locator), cioè l'indirizzo della risorsa, che l'URN (Uniform Resource Name), cioè il nome che la distingue da altre risorse.

<sup>5</sup> Il sistema di autocitazione più diffuso è CiteSeer, che lavora scaricando articoli dal web ed estraendo da questi tutte le citazioni e il contesto delle citazioni stesse. Le funzionalità di CiteSeer sono: il testo pieno, l'indice delle citazioni e la ricerca di parole chiave insieme alla navigazione ai collegamenti citati.

### Abstract

*A citation system allows identification and access to cited resources. It needs both a document's description model and a document's organisation model. For digital documents in the Web, there are two different approaches to citation: one which extends the traditional citation rules and one which uses technologies for linking documents. However these solutions are not completely matching users needs. It is necessary to limit the citation system to digital documents collected in trusted repositories, which are able to assure document's authenticity and preservation.*