

Citare Internet: tradizioni da confermare e miti da sfatare

di Riccardo Ridi

Undici anni dopo un mio precedente contributo sulla citazione di documenti disponibili in Internet [1], stimolato dalla lettura di alcuni recenti articoli sull'argomento [2-5] e da alcuni seminari sulle citazioni bibliografiche che ho tenuto a Venezia fra il 2004 e il 2006 presso Ca' Foscari, Fondazione Querini Stampalia e ASAC, mi sono domandato se fosse già trascorsa una quantità di tempo sufficiente per rivisitare l'argomento con maggiore prospettiva storica su uno strumento di comunicazione ormai non più sperimentale e per verificare se certe acquisizioni si fossero nel frattempo stabilizzate.

La risposta è sostanzialmente positiva in entrambi i casi, nel senso che si può ormai dare per assodato che fonti provenienti da Internet, se affidabili¹, possono e devono essere utilizzate proficuamente e citate appropriatamente in ogni contesto bibliografico, dal servizio giornalistico alla tesi di laurea fino al saggio scientifico, e che le tesi di fondo enunciate o comunque implicite in quel mio primo lavoro [1] continuano, a mio avviso, ad essere tuttora valide e applicabili. Per comodità sia mia che del lettore sintetizzo e aggiorno qui tali tesi, rinviando al testo dell'epoca e al successivo, scritto in collaborazione con Fabio Metitieri [6, p. 70-71], per le rispettive argomentazioni.

A. Nonostante tutti i discorsi su URN (Uniform Resource Name) e URC (Uniform Resource Characteristics), l'unico URI (Uniform Resource Indicator) realmente esistente e universalmente utilizzato era (nel 1995) e resta (nel 2006) l'ormai onnipresente URL (Uniform Resource Locator), sempre più spesso utilizzato – erroneamente – senza l'iniziale *http://*, visto il clamoroso successo del Web rispetto agli altri protocolli di Internet. L'unico altro identificatore di risorse digitali sviluppatosi nel frattempo che appare oggi come un possibile futuro concorrente o erede dell'URL nell'ambito delle citazioni bibliografiche è il DOI (Digital Object Identifier) [7, 8].

RICCARDO RIDI, Università "Ca' Foscari", Dipartimento di studi storici, calle del piovàn, San Marco 2546, 30124 Venezia, e-mail ridi@unive.it. Gli URL sono stati controllati l'ultima volta il 24 agosto 2006. L'autore ringrazia Elena Boretti, Francesco Dell'Orso, Juliana Mazzocchi, Carlo Revelli e Michele Santoro per informazioni e suggerimenti.

¹ Per una sintesi dei principali criteri per la valutazione di un documento reperito in Internet si veda [6, p. 74-78].

B. Per quanto comodo e di successo, l'URL (e lo stesso vale e varrà per il DOI o per analoghi identificatori futuri) non può assolutamente sostituire tutti gli altri elementi della citazione bibliografica, ma va inteso come un elemento aggiuntivo, oltretutto al confine fra una funzione più prettamente bibliografica e una di localizzazione catalografica, tanto da potersi addirittura ipotizzare un futuro in cui, grazie alla diffusione di repertori costantemente aggiornati e di software per il *reference linking*, esso potrà essere omesso nelle citazioni.

C. Nel contesto di un *medium* dinamico e ancora non sottoposto in tutti i paesi al deposito legale o comunque a efficaci forme di conservazione sistematica e corrente, l'elemento citazionale cronologico diventa ancora più importante che in ambito cartaceo, tanto da poterlo differenziare – almeno in linea di principio – in addirittura sei date:

- creazione iniziale del contenuto intellettuale;
- ultimo aggiornamento del contenuto intellettuale;
- creazione iniziale della attuale veste editoriale;
- prima messa in linea della attuale veste editoriale;
- ultimo aggiornamento degli aspetti puramente redazionali;
- ultima consultazione effettuata da chi cita².

Per collocare temporalmente nel modo migliore un documento reperito *on line* sarebbe meglio citare almeno tre di tali date, scelte di volta in volta nel modo più appropriato, ma spesso si è costretti ad accontentarsi di un paio di quelle più facili da individuare, ovvero quella dell'ultima consultazione e – a seconda della tipologia di risorsa – quella di un ultimo aggiornamento che spesso non si sa se attribuire al contenuto o alla forma del documento stesso oppure quella della pubblicazione a stampa corrispondente.

D. Nel citare documenti recuperabili in Internet è raccomandabile seguire il più possibile sia l'individuazione che la successione dei canonici elementi citazionali, entrambe declinabili in tutta la varietà – ben illustrata da Carlo Revelli [9] – degli stili esistenti o sensatamente inventabili, magari col supporto di guide specializzate come quelle repertorate da Stefania Manzi e Alessandro Corsi [10]. Semmai, laddove risultassero inapplicabili, si potranno omettere elementi come il luogo e la paginazione.

E. Analogamente, si eviterà di creare immotivate distinzioni fra un elenco di documenti più tradizionali, denominato *bibliografia*, e un elenco di documenti recuperabili in Internet, denominato nei modi più vari e talvolta bizzarri, come testimoniato da questi esempi³:

Istography. In: Byron C. Mayes. *Internet istographic tools: citing the sites*. In: *National Online Meeting Proceedings 1997*, edited by Martha E. Williams. Medford: Information today, 1997.

Linkgrafia. In: Laura Corti. *I beni culturali e la loro catalogazione*. Milano: Bruno Mondadori, 2003.

² Quest'ultima, in particolare, andrebbe indicata sempre, meglio se dichiarandola una volta per tutte all'inizio o alla fine della bibliografia, senza appesantirne le singole descrizioni.

³ Ulteriori esempi in italiano e in inglese (fra i quali *internetgrafia*, *netografia*, *urlgrafia* e *urlografia*), con una tabella relativa alla loro presenza sul Web sono offerti da Vittorio Bica [11].

Linkografia. In: Ministero dell'istruzione. *Linkografia e-learning*. Osservatorio tecnologico per la scuola, 2004, <http://www.osservatoriotecnologico.it/internet/e-learning/linkografia_e-learning.htm>.

Linkopedia. In: *Manuale per la qualità dei siti web pubblici culturali*. Seconda edizione italiana aggiornata a cura di Fedora Filippi. Roma: Ministero per i beni e le attività culturali, 2005, oppure <<http://www.minervaeurope.org/publications/qualitycriteria-i/indice0512.html>>.

Netgrafia. In: Piero Cavaleri. *Servizi personalizzati online della biblioteca*. «Biblioteche oggi», 19 (2001), n. 7, p. 70-90, oppure <<http://www.bibliotecheoggi.it/2001/20010707001.pdf>>.

Sitografia. In: Alina Renditiso. *La Sezione IFLA sulla competenza informativa al congresso di Buenos Aires 2004*. «Bibliotime», 7 (2004), n. 3, <<http://didattica.spbo.unibo.it/aiber/bibttime/num-vii-3/renditis.htm>>.

Sitografia essenziale. In: Silvia Trani. *Il patrimonio archivistico italiano sul Web: sitografia essenziale*. «Il mondo degli archivi», 12 n.s. (2004), n. 1-2, p. 235-264.

Sitografia orientativa. In: Marilena Maniaci. *Archeologia del manoscritto: metodi, problemi, bibliografia recente*. Roma: Viella, 2002.

Sitoteca. In: Hellmut Riediger – Pietro Schenone. *La sitoteca: biblioteca delle opere di consultazione e degli strumenti per la ricerca terminologica e lessicale in Internet*. Milano: Edigeo, [2000?-], <<http://www.edigeo.it/Sitoteca/sitoteca.php>>.

Webgrafia essenziale. In: Marco Maria Sambo. *Labirinti: da Cnosso ai videogames*. Roma: Castelvecchi, 2004.

Webgraphy. In: Francesco Dell'Orso. *Bibliography formatting software*. 13th edition, July 2006. In: *ESB Forum*, <<http://www.burioni.it/forum/ors-bfs/>>.

Webliografia. In: Luisa Carrada. *Scrivere per il web: webliografia*, 2000. In: *Il mestiere di scrivere*, <<http://www.mestierediscrivere.com/testi/fonti.htm>>. La “ricerca webliografica” è citata invece da Barbara Fiorentini. *SfogliareWeb: un contributo alla promozione del modello Open Access in Italia*. «Bibliotime», 8 (2005), n. 3, <<http://didattica.spbo.unibo.it/bibliotime/num-viii-3/fiorenti.htm>>.

Webliography. In: Dariush Alimohammadi. *Are webliographies still in use?* «The electronic library», 22 (2004), n. 2, p. 154-157.

Webografia. In: Elena Boretti. *Primi elementi di “webografia”*, «Bollettino AIB», 38 (1998), n. 1, p. 29-40, oppure <<http://www.aib.it/aib/boll/1998/98-1-029.htm>>.

Webibliography. In: Alberto Salarelli – Anna Maria Tammaro. *La biblioteca digitale*. Milano: Editrice bibliografica, 2000.

D'altronde non si vede perché termini come *bibliografia*, *documento* e *pubblicazione*, utilizzati per secoli per indicare oggetti anche molto diversi fra loro, non possano accogliere all'interno della propria sfera semantica anche la documentazione digitale diffusa via rete [9, p. 9, 42, 44]. Caso mai, all'interno di una unica *bibliografia* di documenti si potranno distinguere (ma solo se risulterà effettivamente utile e sensato, scegliendo di applicare a tale scopo una delle varie tecniche esistenti per scandire e articolare un elenco del genere) i documenti disponibili su diversi supporti o tramite diverse modalità⁴.

⁴ Ad esempio, nulla osta ad affiancare una *filmografia*, una *discografia* o – perché no – una *webgrafia* di letteratura primaria (ovvero di documenti oggetti di uno studio di storia o teoria del cinema, della musica o del Web) a una *bibliografia* di letteratura secondaria (ovvero di documenti, su qualsiasi supporto, che hanno per oggetto i documenti del primo elenco, o che comunque si sono rivelati utili per studiarli). Tuttavia i due elenchi costituiranno, esplicitamente o implicitamente, le due parti della bibliografia offerta a corredo di tale studio, intesa stavolta in senso più ampio come l'insieme dei documenti per un verso o per l'altro connessi allo studio stesso.

Ammissibile invece, anche se talvolta tecnicamente problematica, è la possibilità di distinguere in elenchi completamente separati non tanto diverse tipologie di documenti, quanto i documenti da una parte e, dall'altra, entità che documenti non sono, come enti o persone, ciascuno seguito dall'indicazione di un URL di riferimento⁵.

La sostanziale continuità fra "citazione tradizionale" e "citazione Internet" emerge anche dalle estese indagini documentate in [2-5, 12-14], che testimoniano sia come – prevedibilmente – gran parte dei documenti provenienti dalla Rete citati in tesi o articoli scientifici non siano più presenti all'URL originario già dopo poco tempo, ma anche – un po' meno prevedibilmente – come gran parte di essi possano comunque essere recuperati anche parecchi anni dopo grazie, da una parte, a strumenti di archiviazione tanto potenti quanto finora sottoutilizzati come *Internet Archive* <<http://www.archive.org>> o *UK Web Archiving Consortium* <<http://www.webarchive.org.uk/>> e, dall'altra, a citazioni complete e accurate, che amplino le possibilità di recupero del documento a un nuovo URL tramite repertori e motori di ricerca.

A conclusioni sostanzialmente analoghe è del resto giunto anche Giovanni Bergamin, presentando al convegno *Le biblioteche per la libertà d'accesso all'informazione*⁶ i risultati di una più ristretta indagine effettuata sulle 35 citazioni incluse nella versione 2.0 del lavoro di Stefania Manzi e Alessandro Corsi [10], risalente al 1997. Ben 26 di esse nel 2006 non funzionavano più, 21 delle quali però sono state recuperate integralmente in *Internet Archive*, che ne avrebbe salvate anche altre 3, se non avesse ricevuto un esplicito diniego da parte dei relativi autori.

D'altronde i dati – spesso ricordati *en passant*, senza né approfondirne il senso né documentarne le fonti – sulla breve vita media delle pagine web (44 giorni secondo Brewster Kahle, fondatore dell'*Internet Archive*, in un suo articolo del 1997 [15]; 70 secondo lo stesso Kahle, ma in una intervista del 1998⁷; 75 per *Alexa Internet*, già braccio commerciale dell'*Internet Archive*, nel 2001; 100 di nuovo per Kahle, ma in un'altra intervista del 2001, e infine di nuovo, guarda caso, 44 nel 2004 secondo il comu-

5 Chiaramente, fornendo un URL, ciò che viene indicato è anche in questo caso, a rigore, un documento. Esso però non serve direttamente come oggetto di studio, ma come mezzo per individuare, localizzare ed eventualmente contattare un ente o una persona utili per lo studio stesso. Certo, strada facendo ci si potrà anche *documentare* su tale ente o persona, ma anche in tal caso il sito può essere visto più come un dispensatore di singoli micro-documenti – quasi come se fosse una biblioteca o un archivio – piuttosto che come un macro-documento esso stesso. La sofisticata e ambigua dialettica contenuto/contenitore in ambiente ipertestuale, d'altronde, può comunque creare talvolta qualche imbarazzo in chi cita, costretto a considerare come unità documentaria autosufficiente ciò che, in altri *media*, considereremmo semplici frammenti, lacerti o, appunto, *pagine*.

6 Organizzato dall'Associazione italiana biblioteche e tenutosi a Bologna il 18 e 19 maggio 2006; programma e abstract delle relazioni a <<http://www.aib.it/aib/congr/co605/co605.htm3>>.

7 «Brewster Kahle pointed out that research at the *Internet Archive* finds half of the Web disappearing every month, even as it doubles every year. 'The mean life of a Web page is about 70 days', he said. See *No Way to Run a Culture*, by Steve Melon, in "Wired News", February 13, 1998. It should be noted that estimates of the mean life of a Web page vary» [16, chapter 3, <<http://bob.nap.edu/html/lc21/ch3.html>>]. L'intervista originale, dopo varie peregrinazioni all'interno del sito della rivista *Wired*, è attualmente disponibile a <<http://www.wired.com/news/culture/0,1284,10301,00.html>>.

nicato stampa di lancio dell'*UK Web Archiving Consortium*⁸) sono in realtà basati su un equivoco di fondo. Infatti le pagine web migrano spesso, sicuramente troppo e troppo rapidamente, ma per quanto riguarda la loro scomparsa definitiva, essa si verifica molto più raramente, tanto che talvolta ci si scontra con problemi sostanzialmente opposti, ovvero l'ibernazione perenne – nota anche come fossilizzazione – di pagine che invece richiederebbero manutenzione e aggiornamento [18], e la immotivata proliferazione di versioni di uno stesso documento, fra le quali non è facile individuare quella più autorevole e aggiornata.

Quindi ciò che sopravvive dai 44 ai 100 giorni, ammesso che tali dati siano effettivamente attendibili, non sono le pagine web, ma i loro URL. Per quanto riguarda in particolare i documenti reperibili in Internet citati nella letteratura scientifica e nelle tesi di laurea, un'indicazione più plausibile sulla persistenza media dei relativi URL è fornito piuttosto da Carmine Sellitto [4], che sulla base di una ricerca originale e di altre precedenti indagini su questo tipo di materiali indica in un lasso di tempo che va dai 2 ai 4,8 anni (anche in base alla disciplina) la *half-life* degli URL, intesa come il tempo necessario perché metà degli indirizzi contenuti in una bibliografia risulti non più funzionante.

Un'altra considerazione che emerge dalle indagini citate è che non tutti i documenti scompaiono (o, meglio, si nascondono) alla stessa velocità. In media le risorse dotate di caratteristiche che ne garantiscano l'affidabilità rispetto al contenuto informativo sono anche quelle più affidabili dal punto di vista della permanenza al medesimo URL o quantomeno della recuperabilità presso un altro indirizzo. Quindi, incrementando le capacità dei "cimatori" (cosa che varrebbe la pena di fare soprattutto per gli studenti, con uno sforzo congiunto di insegnanti e bibliotecari [3, 13]) di riconoscere le fonti informative *on line* più autorevoli, si aumenterebbero di conseguenza anche la stabilità e la recuperabilità delle citazioni stesse. Un motivo in più, semmai ne mancassero, per spingere sul pedale della *information literacy* nelle biblioteche italiane, soprattutto quelle universitarie e – quando esistono – quelle scolastiche.

RIFERIMENTI BIBLIOGRAFICI

[1] Riccardo Ridi. *Citare Internet*. «Bollettino AIB», 35 (1995), n. 2, p. 211-220 oppure, in versione aggiornata al 27 agosto 1997, <<http://www.aib.it/aib/boll/1995/95-2-211.htm>>.

[2] Mary F. Casserly – James E. Bird. *Web citation availability: analysis and implications for scholarship*. «College & research libraries», 64 (2003), n. 4, p. 300-317.

[3] Joy Thomas. *Web site demise and graduate research: persistence of web pages cited in social work theses*. «Behavioral & social sciences librarian», 22 (2004), n. 2, p. 67-77.

[4] Carmine Sellitto. *A study of missing Web-cites in scholarly articles: towards an evaluation framework*. «Journal of information science», 30 (2004), n. 6, p. 484-495.

⁸ Disponibile a <<http://info.webarchive.org.uk/pressrelease21-06-04.html>>, privo però di link o riferimenti a un eventuale nuovo studio originale che giustifichi il valore dichiarato. Resta dunque il dubbio, piuttosto borgesiano, di trovarci ancora una volta di fronte alla perdurante eco di stime variabili, ma tutte riconducibili allo stesso autore. Per i riferimenti bibliografici relativi alle altre stime citate si veda il paragrafo 2.3.2 del contributo di Michael Day [17].

- [5] Carmine Sellitto. *The impact of impermanent Web-located citations: a study of 123 scholarly conference publications*. «Journal of the American society for information science & technology», 56 (2005), n. 7, p. 695-703.
- [6] Fabio Metitieri – Riccardo Ridi. *Biblioteche in rete: istruzioni per l'uso*. Terza edizione riveduta e aggiornata. Roma-Bari: Laterza, 2005, oppure <<http://www.laterza.it/bibliotecheinrete>>.
- [7] Giuseppe Vitiello. *L'identificazione degli identificatori*. «Biblioteche oggi», 22 (2004), n. 2, p. 67-80, oppure <<http://www.bibliotecheoggi.it/2004/20040206701.pdf>>.
- [8] Mario Sebastiani. *Identificatori persistenti per gli oggetti digitali*. «DigItalia», 0 (2005), p. 62-82, oppure <<http://digitalia.sbn.it/genera.jsp?s=26>>.
- [9] Carlo Revelli. *Citazione bibliografica*. Roma: Associazione italiana biblioteche, 2002.
- [10] Stefania Manzi – Alessandro Corsi. *Citare Internet: un repertorio di risorse in rete*. Versione 3.0, 23 giugno 2006. In: *ESB Forum*, <<http://www.burioni.it/forum/citare.htm>>.
- [11] Vittorio Bica. *Linkografia*, messaggio inviato il 28 novembre 2004 al forum *Lingua italiana e accessibilità*. In: *Diodati.org*, <<http://forum.diodati.org/messaggi.asp?f=6&id=312>>.
- [12] Carol Anne Germain. *URLs: Uniform Resource Locators or Unreliable Resource Locators*. «College & research libraries», 61 (2000), n. 4, p. 359-365.
- [13] Philip M. Davis – Suzanne A. Cohen. *The effect of the Web on undergraduate citation behavior 1996-1999*. «Journal of the American society for information science & technology», 52 (2001), n. 4, p. 309-314.
- [14] Philip M. Davis. *The effect of the Web on undergraduate citation behavior: a 2000 update*. «College & research libraries», 63 (2002), n. 1, p. 53-60.
- [15] Brewster Kahle. *Preserving the Internet*. «Scientific American», 276 (1997), n. 3, p. 72-74.
- [16] Committee on an Information Technology Strategy for the Library of Congress, Computer Science and Telecommunications Board, National Research Council. *LC21: a digital strategy for the Library of Congress*. Washington: The National Academies Press, 2000. Parzialmente disponibile online gratuitamente a <<http://www.nap.edu/catalog/9940.html>>.
- [17] Michael Day. *Collecting and preserving the World Wide Web: a feasibility study undertaken for the JISC and Wellcome Trust*. Version 1.0, 25 February 2003. London: The Joint Information Systems Committee, <http://www.jisc.ac.uk/uploaded_documents/archiving_feasibility.pdf>.
- [18] Ziv Bar-Yossef – Andrei Z. Broder – Ravi Kumar – Andrew Tomkins. *Sic transit gloria telae: towards an understanding of the web's decay*. In: *Proceedings of the 13th international conference on World Wide Web*. New York: ACM Press, 2004, p. 328-337, oppure <<http://www2004.org/proceedings/docs/1p328.pdf>>.

Quoting Internet: traditions to confirm and myths to destroy

by Riccardo Ridi

Documentary sources from Internet, if trustworthy, can and must be used and quoted without qualms in every bibliographical context. The URL (Uniform Resource Locator) indication cannot however replace all the other usual elements of a bibliographical quote, which should be identified and treated while continuing as far as possible to follow tradition, with special attention to chronological elements.

In the same way, what should certainly be avoided is the creation of unmotivated distinctions between a list of more traditional documents, called *bibliography*, and a list of documents that can be recovered from Internet and which are called in the most varied and sometimes bizarre ways.

The basic continuity between “traditional quotation” and “Internet quotation” emerges also from some recent investigations that testify how a large amount of the documents from the Net quoted in theses or scientific articles are no longer present in the original URL a short time later. It is also clear how a large number of them can however be recovered even many years later thanks, on the one hand, to archive tools such as *Internet Archive* and, on the other, to complete and accurate quotations.

After all certain alarmist data on the average short life of web pages is based on a fundamental misunderstanding, because actually what survives from 44 to 100 days are not the web pages but their URLs. As regards the subsets of the documents quoted in scientific literature and in graduation theses, a more plausible estimate on the average life of the relative URL wavers between 2 and 4.8 years.

Another consideration that emerges from the investigations is that the most trustworthy sources as regards information content are also those most stable from the point of view of their permanence in the same URL or at least of their possibility of being recovered from another address. Therefore, by increasing the capacity of the “quoters” to recognize the more authoritative information sources on-line, the possibility of recovery of the quotations themselves would also be increased.

RICCARDO RIDI, Università “Ca’ Foscari”, Dipartimento di studi storici, calle del piovano, San Marco 2546, 30124 Venezia, e-mail ridi@unive.it.

The author thanks Elena Boretti, Francesco Dell’Orso, Juliana Mazzocchi, Carlo Revelli and Michele Santoro for their useful suggestions.