

Metadati descrittivi per il Web Archiving

**Consigli da OCLC Research Library
Partnership Web Archiving
Metadata Working Group**

Jackie Dooley e Kate Bowers

RIEPILOGO GENERALE

“I modi e gli strumenti per condurre la ricerche accademiche stanno subendo drastici cambiamenti, con evidenti conseguenze sul piano della comunicazione della ricerca e in ultima analisi sull’intera letteratura accademica ...” Così inizia l’autorevole report di OCLC Research *The Evolving Scholarly Record*, il quale passa in rassegna la rapida evoluzione del patrimonio archivistico e bibliotecario dal materiale a stampa al digitale, e le importanti conseguenze per le nostre collezioni e servizi. Il contenuto di molte risorse digitali è intrinsecamente mutevole e sempre più relegato a una registrazione e disseminazione esclusivamente online.

Questo rende imperativo preservare i contenuti del web in modo tempestivo, soprattutto se abbiamo intenzione di mantenere l’integrità e la continuità della produzione culturale, storica e accademica. Le biblioteche e gli archivi hanno una lunga esperienza nell’affrontare le sfide dettate dalla conservazione dei supporti di breve durata, ma la longevità dei siti web è ancora più effimera. I siti tendono a venire aggiornati, estesi o riformulati ripetutamente nel tempo, e qualunque sito potrebbe scomparire senza preavviso. Se non verranno archiviati periodicamente, una percentuale significativa dei contenuti web potrà semplicemente smettere di esistere.

Parte del lavoro prende spunto da due recenti sondaggi, uno sui fruitori dei contenuti web archiviati e l’altro relativo ai professionisti che si occupano di web archiving. Entrambi i sondaggi dimostrano come l’assenza di un approccio comune alla creazione di metadati sia la sfida più sentita da tutta la comunità di settore.

Di tutta risposta, OCLC Research ha costituito il Web Archiving Metadata Working Group per sviluppare delle raccomandazioni in materia di metadati descrittivi. I criteri adottati sono stati ponderati sulle peculiari caratteristiche dei siti web archiviati, con un occhio di riguardo nell’aiutare gli istituti della cultura a migliorare la consistenza e l’efficacia delle loro pratiche di metadattazione in questa nuova era informatica.

Abbiamo iniziato con un’analisi della letteratura di settore per una maggiore comprensione delle informazioni di maggior rilievo per gli utenti. Un aspetto è legato al contesto nel quale un sito web o una collezione sono stati archiviati. Questo è di vitale importanza per i fruitori al fine di capire in che modo si possano utilizzare tali contenuti. In risposta alle esigenze espresse dai professionisti dei metadati, abbiamo fornito un ponte tra approccio bibliografico e approccio archivistico alla descrizione - un metodo neutrale che cresce in importanza man mano che nuovi tipi di contenuti digitali permeano le nostre collezioni.

Le biblioteche e gli archivi creano record di metadati sia per siti web live sia archiviati, usando metodologie archivistiche e bibliografiche per descrivere sia i singoli siti che collezioni archiviate. Ogni comunità applica i propri standard e le relative pratiche di catalogazione e descrizione, sebbene talvolta vengano messe in essere anche tecniche ibride. Abbiamo definito un set base di data element con note d’uso per guidare la preparazione dei contenuti di dati. Tali note d’uso possono essere utilizzate in concerto con gli standard esistenti che presentano data element molto più granulari. È un approccio scalabile, neutrale rispetto alla comunità e agli standard, che non richiede descrizioni approfondite o ripetute modifiche ai record nel corso del tempo.

I possibili casi d'uso includono *Ricercatori* che costruiscono archivi personali di siti web per agevolare le proprie ricerche; *biblioteche* che utilizzano Resource Description and Access (RDA)/MARC e che sono in cerca di linee guida specifiche sui data element e sui contenuti più pertinenti per la descrizione dei contenuti web; *archivi* che mappano i loro record MARC 21- basandosi su Describing Archives: A Content Standard (DACS) e/o su Encoded Archival Description (EAD) - verso una struttura più elementare come Archive-it o un repository digitale; *repository digitali* che codificano i metadati per i contenuti web in Metadata Object Description Schema (MODS) senza riferimento ad alcuno standard; e gli utenti di *Archive-it* che cercano dei riferimenti per creare contenuti seguendo il Dublin Core.

INTRODUZIONE

“Il Web Archiving si colloca alla frontiera dell’acquisizione e della conservazione del nostro patrimonio culturale e storico.” Sara Day Thomson, 2016.*

“È ben più facile recuperare un film del 1924 che un sito web del 1994.”
M.S. Ankerson, 2011.†

“I modi e gli strumenti per condurre le ricerche accademiche stanno subendo drastici cambiamenti, con evidenti conseguenze sul piano della comunicazione della ricerca e in ultima analisi sull’intera letteratura accademica - ovvero il resoconto accurato dei passati sforzi della ricerca”. Così inizia l’autorevole report di OCLC Research *The Evolving Scholarly Record*, il quale passa in rassegna la rapida evoluzione del patrimonio archivistico e bibliotecario dal materiale a stampa al digitale, e le importanti conseguenze per le nostre collezioni e servizi. Il contenuto di molte risorse digitali è intrinsecamente mutevole e sempre più relegato a una registrazione e disseminazione esclusivamente online.

Questo rende imperativo preservare i contenuti del web in modo tempestivo se abbiamo intenzione di mantenere l’integrità e la continuità della produzione culturale, storica e accademica. Le biblioteche e gli archivi hanno una lunga esperienza nell’affrontare le sfide per la conservazione dei supporti di breve durata, ma la longevità dei siti web è ancora più effimera: i siti tendono a venire aggiornati, estesi o riformulati ripetutamente nel tempo, e qualunque sito potrebbe scomparire senza preavviso. Senza archiviazione periodica, una percentuale significativa dei contenuti web potrebbe semplicemente smettere di esistere

Le sfide per la conservazione sono difficoltose, così come lo sono quelle relative ai metadati descrittivi per rendere una risorsa web rintracciabile. *Metadati descrittivi per il Web Archiving* è stato pensato in risposta a queste necessità. In parte il lavoro prende spunto da due recenti sondaggi, uno sui fruitori dei contenuti web archiviati e l’altro relativo ai professionisti che si occupano di web archiving - entrambi i sondaggi dimostrano come l’assenza di un approccio comune alla creazione di metadati sia la sfida più sentita da tutta la comunità di settore.^{3,4}

Di tutta risposta, OCLC Research ha costituito il Web Archiving Metadata Working Group (WAM) per sviluppare delle linee guida per i metadati descrittivi e per facilitare la ricercabilità dei contenuti web archiviati.⁵ Il nostro approccio si è adattato alle caratteristiche uniche dei siti web archiviati, con un occhio di riguardo nell’aiutare gli istituti della cultura a migliorare la consistenza e l’efficacia delle loro pratiche di metadattazione in questa nuova era informatica.

* Thomson, Sara Day. 2016. “Surveying the Domain: Three Days with the Web Archiving Team.” *The British Library Web Archive Blog*. Posted 14 September 2016. <http://blogs.bl.uk/webarchive/2016/09/surveying-the-domain-three-days-with-the-web-archiving-team-.html>.

† Ankerson, Megan Sagnar. 2011. “Writing Web Histories with an Eye on the Analog Past.” *New Media & Society*. doi:10.1177/1461444811414834.

Due definizioni di base aiutano a inquadrare il contesto di questo lavoro: il *Web archiving* è "il processo di raccolta e messa a disposizione perpetua dei contenuti web", mentre per *archivi web* abbiamo due accezioni: "1. Copie conservate di contenuti web live per la conservazione e l'accesso permanente 2. un'organizzazione che si dedica principalmente alla raccolta e alla conservazione di contenuti web".⁷

Il gruppo di ricerca ha riconosciuto l'importanza di ottenere una chiara comprensione dei bisogni dei fruitori dei contenuti web archiviati, e lo ha tenuto in considerazione durante tutte le fasi del proprio lavoro. Analizzando la letteratura di settore abbiamo acquisito maggiore consapevolezza delle informazioni di maggior rilievo per gli utenti. Un aspetto di grande importanza è il contesto nel quale un sito web o una collezione vengono archiviati, ciò è fondamentale per i fruitori al fine di capire in che modo si possano utilizzare tali contenuti. In risposta alle esigenze espresse dai professionisti dei metadati, abbiamo fornito un punto d'incontro tra approccio bibliografico e approccio archivistico alla descrizione.

Per assicurarci di non duplicare il lavoro pianificato da altri, prima di intraprendere questo progetto, ci siamo consultati con l'International Internet Preservation Consortium⁸, con la Society of American Archivists Web Archiving Section⁹, e con il programma Archive-It di Internet Archive¹⁰. Ciascuno di loro ci ha incoraggiato a procedere. Ci siamo anche impegnati a instaurare un dialogo costante con la comunità di settore, sia per riferire sui progressi compiuti sia per sollecitare dei feedback. Nel 2016 abbiamo pubblicato un rapporto sui lavori in corso e lo abbiamo presentato a numerose conferenze. Abbiamo anche pubblicato gli aggiornamenti su listservs e sul blog di OCLC Research, hangingtogether.org.

I risultati del nostro lavoro si trovano in questa pubblicazione, e in altre due: *Descriptive Metadata of Web Archiving: Literature Review of User Needs*¹³ and *Descriptive Metadata of Archiving: Review of Harvesting Tools*.¹⁴

RIFERIMENTI BIBLIOGRAFICI

1. Lavoie, Brian, Eric Childress, Ricky Erway, Ixchel Faniel, Constance Malpas, Jennifer Schaffner, and Titia van der Werf. 2014. *The Evolving Scholarly Record*. Dublin, OH: OCLC Research. <http://www.oclc.org/research/publications/library/2014/oclcresearch-evolvingscholarly-record-2014.pdf>.
2. The ideas formulated in the 2014 report are elaborated in: Lavoie, Brian, and Constance Malpas. 2015. *Stewardship of the Evolving Scholarly Record: From the Invisible Hand to Conscious Coordination*. Dublin, OH: OCLC Research. <http://www.oclc.org/content/dam/research/publications/2015/oclcresearch-esrstewardship-2015-a4.pdf>.
3. Ricky Erway. 2015. "Thoughts from Partner Staff about Web Archiving" *hangingtogether.org* (blog). Posted 29 October 2015. <http://hangingtogether.org/?p=5450>.
4. A research team led by Matthew Weber at Rutgers University surveyed users of web archives in the winter of 2016. They expect to publish their data soon.
5. OCLC Research Library Partnership "Web Archiving Metadata Working Group." Last updated 10 March 2016. <http://oc.lc/wam>.
6. "Web Archiving." 2017. Word of the Week, Society of American Archivists. <http://us3.campaign-archive2.com/?u=56c4cfbec1ee5b2a284e7e9d6&id=40edf162c4>.
7. See note 6.
8. <http://netpreserve.org/>.
9. <https://www2.archivists.org/groups/web-archiving-section>.
10. <https://archive-it.org/>.
11. Dooley, Jackie M., Karen Stoll Farrell, Tammi Kim, and Jessica Venlet. 2017. "Developing Web Archiving Metadata Best Practices to Meet User Needs." *Journal of Western Archives* 8:2. <http://digitalcommons.usu.edu/westernarchives/vol8/iss2/5/>.
12. A sample blog post: Dooley, Jackie. 2017. "Best Practices for Web Archiving Metadata: Watch this Space!" *hangingtogether* (blog). Posted 5 April 2017. <http://hangingtogether.org/?p=5918>.
13. Samouelian, Mary, and Jackie Dooley. 2018. *Descriptive Metadata for Web Archiving: Review of Harvesting Tools*. Dublin, OH: OCLC Research. doi:10.25333/C37H0T.
14. Venlet, Jessica, Karen Stoll Farrell, Tammy Kim, Allison Jai O'Dell, and Jackie Dooley. 2018. *Descriptive Metadata for Web Archiving: Literature Review of User Needs*. Dublin, OH: OCLC Research. doi:10.25333/C33P7Z.