



AIUCD 2022 | UNIVERSITÀ DEL SALENTO

CULTURE DIGITALI

INTERSEZIONI

FILOSOFIA

ARTI

MEDIA

TESTO

ARTI

FILOSOFIA

CONTENUTI

INTELLIGENZA

PROCEEDINGS

ISBN 9788894253566

Copyright ©2022 AIUCD
Associazione per l'Informatica Umanistica e la Cultura Digitale



Il presente volume e tutti i contributi sono rilasciati sotto licenza Creative Commons Attribution Share-Alike 4.0 International license (CC-BY-SA 4.0). Ogni altro diritto rimane in capo ai singoli autori.



This volume and all contributions are released under the Creative Commons Attribution Share-Alike 4.0 International license (CC-BY-SA 4.0). All other rights retained by the legal owners.

Fabio Ciraci, Giulia Miglietta, Carola Gatto (edd.), AIUCD 2022 - Culture digitali. Intersezioni: filosofia, arti, media. Proceedings della 11ª conferenza nazionale, Lecce, 2022. Fabio Ciraci, Giulia Miglietta, Carola Gatto (edd.), AIUCD 2022 - Digital cultures. Intersections: philosophy, arts, media. Proceedings of the 11th national conference, Lecce, 2022.

Salvo diversa indicazione, ogni link citato era attivo al 21 gennaio 2022. All links have been visited on 21th January 2022, unless otherwise indicated

Si prega di notificare all'editore ogni omissione o errore si riscontri, al fine di provvedere alla rettifica. Please notify the publisher of any omissions or errors found, in order to rectify them. [aiucd.segreteria \[at\] aiucd.org](mailto:aiucd.segreteria@aiucd.org)

I contributi pubblicati nel presente volume hanno ottenuto il parere favorevole da parte di valutatori esperti della materia, attraverso un processo di revisione anonima mediante *double-blind peer review* sotto la responsabilità del Comitato Scientifico di AIUCD 2022.

All the paper published in this volume have received favourable reviews by experts in the field of DH, through an anonymous double-blind peer review process under the responsibility of the AIUCD 2022 Scientific Committee.

Il programma della conferenza AIUCD 2022 è disponibile online all'indirizzo/ The AIUCD 2022 conference program is available online all'apposito indirizzo

<http://aiucd2022.unisalento.it> <http://conference.unisalento.it/ocs/index.php/aiucd2022/index/pages/view/programma>

Comitato Scientifico:

General Chair: Fabio Ciraci (Università del Salento)

Local Chair: Mario Bochicchio (Università del Salento, Università di Bari)

Membri Comitato Scientifico: Marina Buzzoni (Presidentessa AIUCD, Uni. Venezia), Federico Boschetti (Ric. ILC-CNR); Federico Meschini (Uni. Tuscia); Roberto Rosselli Del Turco (Uni Torino); Rachele Sprugnoli (Ass. Ric. Univ. Cattolica); Donato Malerba (Università Bari);

Luca Bandirali, Daniela Castaldo, Francesco Ceraolo, Stefano Cristante, Domenico M. Fazio, Manolita Francesca, Marco Mancarella, Pietro Luigi Iaia, Massimiliano Rossi, Grazia Semeraro, Franco Tommasi, Luigi Patrono (Università del Salento)

Membri del Comitato di programma: Mario Bochicchio (Local Chair), Luca Bandirali, Daniela Castaldo, Marco Mancarella, Pietro Luigi Iaia, Federica Epifani (Responsabile Comitato di Programma), Ilenia Colonna, Patrizia Miggiano; Carola Gatto; Giulia Miglietta; Marco Giannotta; Alessia De Blasi, Isabella Hernandez.

Direttori di Area: Luca Bandirali; Mario Bochicchio; Fabio Ciraci; Roberto Rosselli Del Turco; Marco Mancarella; Grazia Semeraro.

Segreteria del Convegno: Dott.ssa Silvia Gravili

Resp. tecnico: Carlo Tafuro; web design: Dr.ssa Paola D'Amico; comunicazione: Dr.ssa Loredana De Vitis

Enti organizzatori / Organizing institutions:

AIUCD;

Università del Salento: Centro interdipartimentale in Digital Humanities in collaborazione con i corsi di laurea in Filosofia, DAMS, Beni Culturali e Digital Humanities; ISUFI, Scuola Placetelling.

Università degli Studi Aldo Moro, Dipartimento di Informatica

Sponsor

Regione Puglia; Provincia di Lecce; Città di Lecce; CINI – Consorzio Universitario Nazionale per l'Informatica; SFI-Società Filosofica Italiana; AFC - Apulia Film Commission, Teatro Pubblico Pugliese; Argo Software.

Lista dei revisori - List of the reviewers

Agnese Addone; Tommaso Agnoloni; Luca Bandirali; Nicola Barbuti; Andrea Bellandi; Armando Bisogno; Mario Alessandro Bochicchio; Andrea Bolioli; Federico Boschetti; Dominique Brunato; Paolo Buono; Dino Buzzetti; Marina Buzzoni; Luigi Catalani; Francesco Ceraolo; Daniele Chiffi; Simona Chiodo; Fabio Ciotti; Ilenia Colonna; Christian D'Agata; Elisa D'Argenio; Riccardo De Biase; Manuela De Giorgi; Daniela De Leo; Salvatore De Masi; Pierpaolo Del Coco; Angelo Mario Del Grosso; Francesca Di Donato; Giorgio Maria Di Nunzio; Federica Epifani; Daniela Fogli; Claudio Forziati; Greta Franzini; Francesca Frontini; Emiliano Giovannetti; Edmondo Grassi; Fabiana Guernaccini; Barbara Guidi; Pietro Luigi Iaia; Benedetta Iavarone; Fahad Khan; Maurizio Lana; Angelica Lo Duca; Donato Malerba; Marco Mancarella; Tiziana Mancinelli; Chiara Mannari; Valentina Marangi; Cristina Marras; Federico Meschini; Patrizia Miggiano; Giulia Miglietta; Paolo Monella; Giovanni Morrone; Serge Noiret; Deborah Paci; Antonio Pascucci; Enrico Pasini; Luigi Patrono; Igor Pizzirusso; Simone Rebora; Massimiliano Rossi; Daniela Rotelli; Enrica Salvatori; Eva Sassolini; Daria Spampinato; Rachele Sprugnoli; Enrico Terrone; Francesca Tomasi; Francesco Tommasi; Sara Tonelli; Gennaro Vessio; Marco Salvatore Zappatore.

Indice – Table of Contents

Prefazione	I
Sessione Arti 1 – Artemisia Gentileschi	2
La Comédie Virtuelle	4
Climate change & digital cultural impact, the Victoria & Albert Museum	9
La Digitalizzazione per una fruizione del Patrimonio Culturale in sito e da remoto: il caso studio della Pala Gozzi di Tiziano	12
Sessione Testi 1 – Claude Shannon	18
Verso la definizione di criteri per valutare soluzioni di scholarly editing digitale: il caso d’uso GreekSchools	20
HYLAS: A new metrical search tool for Greek and Latin poetry	26
Stylometry and Reader Response. An Experiment with <i>Harry Potter</i> Fanfiction	30
Sessione Intelligenza 1 – Alan M. Turing	35
Analisi e valorizzazione del patrimonio artistico mediante Intelligenza Artificiale	37
Un Oggetto Intelligente IoT per Migliorare le Visite Interattive di Siti di Interesse Culturale	42
Oxoce - Motore di ricerca tematico strutturato	46
Sessione Contenuti 1 – George Boole	49
Funzione ecosistemica e funzione storiografica della narrazione ambientale videoludica	51
Narrazioni mediatiche delle emergenze e processi di costruzione di <i>quest</i> : quali possibili analogie?	
L’incidente del “corrupted blood” in “World of Warcraft”	54
Narrazione e interazione	59
Sessione Testi 2 – Ada Lovelace	61
Web e social media come nuove fonti per la storia	63
Idee, persone, <i>realia</i> : un ambiente digitale per la Via della Seta	68
Visualizzazione del cambiamento d’uso del maschile e femminile nei titoli occupazionali	71
GenderedOntoComedy: Toward a Gendered Representation of Literary Characters in the Dante’s Commedia	76
Sessione Filosofia 1 – Marisa Bellisario	81
Gli indici della prima modernità come strumento storiografico: questioni preliminari metodologiche e pratiche	83
Indici e mappe digitali per l’iter italicum di G. W. Leibniz	86
Ermeneutica digitale del testo filosofico. Problemi e opportunità	91
Human Enhancement e soggetto Post-Umano alla prova delle DH: come le tecnologie digitali ci trasformano	93

Sessione Testi 3 – Grace Murray Hopper	96
Conservazione e fruizione di banche dati letterarie: l'archivio della poesia italiana dell'Otto/Novecento di Giuseppe Savoca	98
«Le varianti della rosa». Per un prototipo di edizione digitale del <i>Nome della rosa</i> : interpretazione, didattica, annotazione	105
Online lexical resources for translators: where do we stand? A (possibly meaningful) case-study	111
 Sessione Filosofia 3 – Gilbert Simondon	 116
Governare le piattaforme. Cinque proposte su pluralismo e polarizzazione online	118
A Taxonomy of Depictive Representations: From Paintings and Sculptures to Virtual Reality	122
Paesaggi dell'incontro mediale on-demand	126
 Sessione Contenuti 2 – Marshall McLuhan	 129
Tra Public e Digital History: la soluzione ibrida dei registri parrocchiali di Monterosso on line	131
Una nuova mappatura digitale per i borghi delle aree interne	138
Intelligenza artificiale e archivi audiovisivi: potenzialità e sfide del progetto "PH-Remix"	141
 Sessione Intelligenza 2 – John von Neumann	 145
Un nuovo approccio per la descrizione e gestione del patrimonio culturale digitale relativo a MAB	147
Sulla funzionalità di un'ontologia della filosofia alto medievale. Il caso dei «Moralia in lob» di Gregorio Magno	151
La Visualizzazione Grafica di Sensi e Relazioni Semantiche di un Lessico Computazionale della Lingua Italiana	155
 Sessione Testi 4 – Hedy Lamarr	 161
Dalla codifica alla fruizione: l'edizione digitale Bellini Digital Correspondence	163
Dante e Petrarca allo (stesso) scrittoio. Per lo sviluppo di un'ontologia di IDP a partire dall'istanza manoscritti di Itinera	169
Il progetto 'epistolarITA' e una proposta di applicazione di algoritmi di prossimità testuale su documenti epistolari italiani (XV-XVII s.)	172
 Sessione Testi 5 – Hélèn Metzger	 177
Visualizing the genetic process of literary works	179
Analisi linguistica e pseudonimizzazione: strumenti e paradigmi	185
RePIM in LOD: semantic technologies to preserve knowledge about Italian secular music and lyric poetry from the 16th-17th centuries	193
 Sessione Filosofia 2 – Giulio Cesare Vanini	 196
Computare o comporre? Riflessioni sul rapporto tra poesia e digitalità alla luce di alcune considerazioni bachelardiane	198
Schemi, ipotesi e algoritmi. Approcci kantiani alla filosofia delle tecnologie digitali	203
Tra chair e empiriquement lo spazio topologico: contributo merleau-pontyano ai sistemi informatici	207

Web e social media come nuove fonti per la storia

Chiara Aldini¹, Stefano Allegrezza², Tommaso Mazzoli³

¹ Università degli Studi di Bologna, Italia, chiara.aldini4@unibo.it

² Università degli Studi di Bologna, Italia, stefano.allegrezza@unibo.it

³ Università degli Studi di Udine, Italia, Tommaso.mazzoli@uniud.it

ABSTRACT

Il contributo intende mettere in evidenza come negli ultimi anni l'interesse verso i temi dell'archiviazione e conservazione del web e dei social media sia cresciuto enormemente, anche perché la consapevolezza dell'importanza di tali "risorse" come fonti privilegiate per ricostruire la storia della nostra epoca è ormai acquisita. Come faranno gli storici del futuro a ricostruire il periodo storico che stiamo vivendo se archivisti e bibliotecari non saranno capaci di archiviare e preservare i siti web e social media di istituzioni, enti pubblici, partiti, associazioni, organi di governo, personaggi politici, personaggi illustri in genere, dato che ormai tutto viene veicolato attraverso tali canali? La fragilità del web, poi, imporrebbe di agire subito ed avviare senza indugio iniziative di "web and social media archiving", pena la scomparsa di quanto è stato pubblicato e reso disponibile sul web negli ultimi anni, ma su questo punto la situazione in Italia – salvo poche eccezioni – appare molto in ritardo rispetto agli altri paesi europei ed enormemente in ritardo rispetto ai paesi dell'area anglosassone. Occorre avviare iniziative di sensibilizzazione su questi temi e di formazione delle competenze e delle professionalità necessarie per condurre progetti di archiviazione e conservazione del web e dei social media.

PAROLE CHIAVE

Web archiving; social media archiving; digital preservation; fonti archivistiche; storia.

INTERVENTO

1. INTRODUZIONE

L'interesse verso i temi della conservazione del web e dei social media è cresciuto notevolmente soprattutto negli ultimi anni, ovvero da quando è emersa sempre più distintamente la consapevolezza che essi costituiranno fonti insostituibili per la comprensione della storia e della civiltà contemporanee. Si pensi, ad esempio, alla pandemia da COVID-19 ancora in corso: le fonti web saranno fondamentali per ricostruire e raccontare gli avvenimenti di questo periodo ([14]) e senza di esse sarà molto difficile ricostruire questi due anni solo a partire dalle fonti tradizionali. È per questo motivo che negli ultimi vent'anni si è molto sviluppato il web archiving, cioè il processo finalizzato alla 'cattura' e conservazione sistematica di porzioni del web a cura di istituzioni della memoria, come archivi e biblioteche nazionali, istituzioni universitarie, fondazioni ([2]). Molti istituti di conservazione si sono attivati con iniziative e progetti di web archiving. Recentemente, si è cominciato anche a sviluppare un ulteriore settore, quello dell'archiviazione e conservazione dei social media, ritenuti anch'essi – non a torto – fonti insostituibili per ricostruire il periodo storico che stiamo vivendo.

2. STATO DELL'ARTE

Le prime riflessioni sul web archiving risalgono alla fine degli anni Novanta, periodo in cui inizia a porsi concretamente il problema della conservazione dei siti web. Nel 1996, sei anni dopo lo sviluppo del World Wide Web ad opera di Tim Berners Lee, prende avvio Internet Archive ([9]), organizzazione senza scopo di lucro con la *mission* di creare una digital library di siti internet e così salvaguardarne e garantirne l'accesso permanente. I fondatori, gli ingegneri statunitensi Brewster Kahle e Bruce Gilliat, 'catturarono' le prime istantanee di pagine web mediante l'utilizzo di un 'crawler', un'applicazione che riusciva a catturare una fotografia fedele e inalterata dei siti presenti sul web. Oggi Internet Archive vanta oltre 25 anni di cronologia web, per un totale di circa 70 petabyte di dati raccolti, resi accessibili tramite la Wayback Machine ([1]), ovvero l'interfaccia pubblica che consente di ricercare e visualizzare le versioni archiviate dei siti web. Inserendo l'URL (Uniform Resource Locator) che si vuole analizzare e selezionando un intervallo di date, è possibile navigare sulla versione archiviata di una pagina o un sito web.

Internet Archive mette a disposizione anche Archive-It, un servizio in abbonamento disponibile dal 2006 che consente alle istituzioni di costruire e preservare raccolte di contenuti nativi digitali. Attraverso l'applicazione web di facile utilizzo, i partner di Archive-It possono raccogliere, metadattare, gestire e generare una copia delle proprie raccolte digitali, le quali vengono archiviate ed ospitate nel data center di Internet Archive e rese accessibili al pubblico con ricerca full-text.

L'importanza di preservare le informazioni digitali è stata riconosciuta nel 2003 anche dall'Organizzazione delle Nazioni Unite per l'educazione, la scienza e la cultura (UNESCO), che nel "Charter on the Preservation of Digital Heritage" inserisce tra i materiali digitali che costituiscono il "digital heritage" anche le pagine web:

«The digital heritage consists of unique resources of human knowledge and expression. It embraces cultural, educational, scientific and administrative resources, as well as technical, legal, medical and other kinds of information created digitally, or converted into digital form from existing analogue resources. Where resources are "born digital", there is no other format but the digital object. Digital materials include texts, databases, still and moving images, audio, graphics, software and web pages, among a wide and growing range of formats. They are frequently ephemeral, and require purposeful production, maintenance and management to be retained. Many of these resources have lasting value and significance, and therefore constitute a heritage that should be protected and preserved for current and future generations. This ever-growing heritage may exist in any language, in any part of the world, and in any area of human knowledge or expression» ([16]).

Nel 2003 viene fondato anche l'International Internet Preservation Consortium (IIPC), che riunisce alcune delle principali biblioteche nazionali e svolge attività di promozione e sviluppo di strumenti, tecniche e standard comuni per la creazione di archivi web internazionali. Attualmente partecipano all'IIPC organizzazioni di oltre 35 paesi, tra cui biblioteche e archivi nazionali, universitari e regionali ([8]).

Sebbene dal 2018 non sia più attiva, occorre citare anche la Internet Memory Foundation (fino al 2010 European Archive Foundation), fondata nel 2004 e coinvolta in progetti di ricerca, finanziati dalla Commissione Europea, volti a migliorare le tecnologie di web crawling, estrazione dati, text mining e conservazione degli archivi web delle istituzioni europee.

Oltre ad iniziative e strumenti per raccogliere e tenere traccia delle risorse sul web, all'interno della comunità internazionale è nata anche l'esigenza di uno specifico formato contenitore che consentisse di archiviare più risorse web in un unico file. Un grosso passo avanti in questa direzione è stata la pubblicazione, nel 2009, dello standard ISO 28500 (versione corrente aggiornata ISO 28500:2017) che ha definito il formato WARC (Web ARChive) – una revisione del formato ARC File Format usato inizialmente da Internet Archive per archiviare le catture del web – che oggi rappresenta il formato standard per gli archivi web ([10]), insieme al formato WACZ (Web Archive Collection Zipped) recentemente proposto ([5]). Nel 2013 viene pubblicato anche lo standard ISO/TR 14873:2013 che definisce principi, metodi e standard di qualità per le istituzioni culturali che si occupano di web archiving ([11]).

Dagli anni Novanta sono numerose le iniziative di archiviazione del web promosse in ambito internazionale. PANDORA, avviato nel 1996 ad opera della National Libraries of Australia, è stato il primo progetto di web archiving sviluppato da un'istituzione pubblica.

Molto attiva l'area del Nord Europa, con i progetti di Svezia (Kulturarw3, 1996), Norvegia (2001), Islanda (2004), Danimarca (Netarkivet, 2005). Il progetto americano della Library of Congress prende avvio nel 2000, quello della Bibliothèque Nationale de France nel 2006. Nel Regno Unito si occupano di web archiving sia i National Archives che le biblioteche incaricate del deposito legale, costituendo così un modello di riferimento internazionale. L'elenco più esaustivo ed aggiornato relativo alle iniziative internazionali di archiviazione del web è stato redatto sulla base dell'indagine condotta dal team di Arquivo.pt, l'archivio web del Portogallo. I risultati, che sono resi disponibili su Wikipedia nella pagina "List of web archiving initiatives" ([13]) evidenziano non solo i progetti ed i relativi paesi di appartenenza, ma anche le tecnologie di web archiving utilizzate ed il personale dedicato, distinguendo anche tra incarichi full-time o part-time.

Nel 2013 la Digital Preservation Coalition ha pubblicato un report specifico sul web archiving nella serie delle DPC Technology Watch Publications ([18]). Il rapporto discute le questioni chiave affrontate dalle organizzazioni impegnate in iniziative di archiviazione del web ed offre una panoramica dei principali software e strumenti attualmente disponibili.

In questo quadro di grandi sforzi a livello internazionale nel tentativo di trovare le strategie per preservare una risorsa che è di per sé estremamente effimera, l'Italia si distingue per l'enorme ritardo rispetto agli altri paesi, anche europei. L'unica iniziativa a livello nazionale meritevole di menzione è stata avviata nel 2018 dalla Biblioteca Nazionale Centrale di Firenze con il progetto di raccolta e archiviazione di siti web di 'interesse culturale' per la storia e la cultura italiana, secondo i principi della legge nazionale sul deposito legale (L. 106/2004 e suo Regolamento attuativo D.P.R. 252/2006) ([15]). La Biblioteca si fa carico, oltre alla raccolta, anche dell'organizzazione e della metadatozione "manuale" dei siti archiviati, avvalendosi per l'accesso e la conservazione della piattaforma Archive-it di Internet Archive. Il deposito legale dei documenti diffusi tramite rete informatica, tuttavia, non è ancora obbligatorio perché lo stesso D.P.R. 252/2006 all'art. 37 prevedeva che il deposito di tali documenti fosse subordinato alla redazione di uno specifico regolamento tecnico che non è ancora stato emanato. Pertanto, l'adesione al programma è su base volontaria da parte dei gestori dei siti, i quali possono

manifestare il proprio interesse compilando il form online disponibile sul sito della Biblioteca. La raccolta si presenta quindi a tutt'oggi molto parziale e frammentata rispetto alla produzione di interesse culturale presente sul web italiano. La scarsa sensibilità sul tema del web archiving e l'assenza di uno specifico quadro normativo rendono difficoltosa l'attuazione di strategie nazionali condivise, laddove un'azione sinergica sarebbe necessaria vista la rapida evoluzione del web e l'enorme quantità di risorse culturali che vi trovano sede e che rischiano di andare irrimediabilmente perdute, stante la rapida evoluzione del web e la sua 'fragilità' ([12]).

3. APPLICAZIONI E STRUMENTI

Per quanto auspicabile, l'individuazione di una strategia condivisa a livello nazionale riguardo la conservazione e la salvaguardia – anche parziale – dei contenuti del web è un traguardo ancora molto lontano da raggiungere.

Sembra quindi quanto mai opportuno che le singole istituzioni, gli enti di ricerca e forse anche i privati cittadini provino a dotarsi di sistemi autonomi di web archiving che possano sopperire a questa mancanza. A questo proposito occorre rilevare che il panorama delle applicazioni software che consentono di archiviare un sito internet è ancora molto frammentato. Vi sono soluzioni pensate per scaricare e consultare siti web con un limitato numero di pagine che possono essere utilizzate da singoli utenti o da gruppi di lavoro ristretti; altre che permettono processi di *harvesting* molto più strutturati e complessi ma che richiedono competenze e attrezzature informatiche decisamente più articolate.

Nel nostro caso, senza avere la pretesa di analizzare tutte le soluzioni presenti sul mercato – che sono numerose e in costante aumento – ci è parso importante segnalare due strumenti software particolarmente interessanti perché, dall'analisi condotta, si sono rivelati particolarmente efficaci pur rimando semplici da utilizzare, e, quindi, costituiscono un buon punto di partenza per un progetto di salvaguardia dei siti web e possono essere utilizzati anche da coloro che non hanno particolari competenze informatiche,

Il software più conosciuto è sicuramente HTTrack, un'applicazione *open source* per il mirroring di siti web e la loro navigazione offline ([7]). Il programma principale si esegue solo con istruzioni testuali dalla riga di comando dei principali sistemi operativi come Linux, Windows e Mac. Esiste però la possibilità di interagire con il software anche grazie ad un'interfaccia grafica predisposta per Windows (WinHTTrack) e per Linux (WebHTTrack) che ne rende sicuramente più semplice l'utilizzo. Questo programma permette di scaricare un intero sito internet sul proprio computer in una cartella a nostra scelta ricostruendone l'intera struttura. Vengono quindi memorizzati il codice, le immagini ed ogni altro tipo di file dal server al computer locale. È importante sottolineare che il software adatta la struttura originale dei link relativi al sito in modo da poterli comodamente navigare in off-line. Vengono supportati i protocolli HTTP e FTP ma non HTTPS. Dopo che il programma avrà eseguito la copia dall'indirizzo web selezionato sarà sufficiente aprire una qualsiasi pagina del sito nel proprio browser per poter navigare esattamente come se si fosse on-line. Il software permette inoltre di configurare numerose opzioni per limitare o estendere la raccolta di base e per controllare il tipo e le caratteristiche dei file da scaricare sul proprio computer. Per ora non è prevista la possibilità di salvare il sito in formato WARC anche se in rete esiste un tool di conversione, *httrack2warc* ([6]), dai risultati ancora non del tutto affidabili. HTTrack è scaricabile dal sito web ufficiale insieme ad una guida dettagliata al tipo di installazione che si intende eseguire e un manuale d'uso per configurare il lavoro di salvataggio del sito web. Purtroppo l'ultima versione del programma è stata rilasciata più di quattro anni fa e ciò fa presupporre che il progetto sia stato abbandonato e quindi non più supportato da nuovi aggiornamenti nel prossimo futuro. Il secondo software che vogliamo segnalare è Cyotek WebCopy ([4]) uno strumento gratuito ma non *open source* che consente di scaricare automaticamente il contenuto di un sito web sul proprio dispositivo locale. Come HTTrack anche WebCopy esegue la scansione del sito web specificato e ne scarica il contenuto: i collegamenti a risorse come fogli di stile, immagini e altre pagine del sito verranno automaticamente rimappati in modo da corrispondere al proprio percorso locale. Utilizzando il pannello di configurazione si potranno definire quali parti di un sito web verranno copiate, permettendo, ad esempio, di scaricare solo le immagini piuttosto che l'intero contenuto. L'interfaccia d'uso è molto completa e sicuramente più facile da utilizzare rispetto HTTrack. Tra le opzioni principali è presente il comando che consente di effettuare una scansione completa del sito prima di iniziare a scaricarlo. Ciò permette anche la creazione della mappa del sito stesso utile ad individuarne la struttura per selezionare zone di particolare interesse. WebCopy riesce anche a scaricare il contenuto di aree protette da password avendo cura di inserire le credenziali richieste in fase di avvio della raccolta. Anche in questo caso non abbiamo la possibilità di salvare il lavoro di copia nel formato WARC. Il software è disponibile sul sito del produttore ma esiste solo la versione per il sistema operativo Windows (dalla versione 7 in poi). Sullo stesso sito è inoltre presente una corposa documentazione di supporto sia per l'installazione che per il suo utilizzo con diversi esempi pratici. L'ultima versione stabile è stata rilasciata alla fine di marzo 2021 ma la presenza di altre versioni in fase di test fa pensare che la casa produttrice sia intenzionata a seguire e migliorare il suo prodotto anche in futuro.

I software qui brevemente presentati sono solo un esempio che dimostra come le pratiche di web archiving possano essere implementate anche all'interno di progetti d'archiviazione di portata limitata e, volendo, anche di tipo personale. Tuttavia, per progetti di più ampio respiro, come quelli portati avanti dalle biblioteche nazionali di alcuni Stati europei e dell'area anglosassone, questi strumenti non sono più sufficienti ed è necessario mettere in campo risorse – non solo tecniche ed economiche, ma anche umane – di dimensioni molto più ampie.

4. CONCLUSIONI

Se è vero che il tema dell'archiviazione e conservazione del web e dei social media sta acquisendo una rilevanza sempre maggiore, è altrettanto vero che ad oggi le figure professionali capaci di condurre progetti in questo ambito sono poche se non addirittura quasi del tutto assenti ([3]), salvo casi eccezionali e certamente meritevoli di segnalazione – come quello già citato della Biblioteca Nazionale Centrale di Firenze. Ciò è dovuto, da una parte, allo scarso interesse che fino ad oggi è stato riservato a questi temi e, dall'altra, alla mancanza di percorsi formativi che sarebbero invece estremamente importanti anche in considerazione delle difficoltà – non solo di tipo tecnico ma anche economico ed organizzativo – che devono essere superate per portare a termine progetti di questa natura. Per superare queste difficoltà, dal 6 al 10 settembre 2021 presso l'Università degli Studi di Bologna si è tenuta la prima edizione della Summer school in “Web and social media archiving and preservation”, con l'intento di:

«offrire una formazione di alto livello sui temi emergenti dell'archiviazione e conservazione dei siti web e dei social media, che rappresentano una nuova e diversificata tipologia di materiale la cui conservazione è imprescindibile per tutta una serie di ambiti scientifici (si pensi alla ricerca storica, sociologica, antropologica, etc.) ai fini della futura ricostruzione dell'attuale civiltà. La Summer school intende anche fornire le conoscenze e le competenze necessarie per favorire lo sviluppo di nuove professionalità ed avviare nuovi percorsi lavorativi da parte dei discenti interessati» ([17]).

La prima edizione della Summer school ha visto la partecipazione di quasi 40 partecipanti tra professionisti dei beni culturali (archivisti, bibliotecari, operatori museali), informatici, funzionari di enti pubblici ed aziende private, studenti/dottorandi in Library and information science e in Digital humanities oltre che persone interessate a vario titolo alle questioni legate all'archiviazione e alla conservazione dei siti web, dei blog e dei social media. Con questa iniziativa si è voluto, da una parte, avviare un percorso di sensibilizzazione sull'importanza della conservazione del web e dei social media, dall'altra, formare i professionisti che saranno in grado di gestire i processi di archiviazione e conservazione almeno di quella parte del web e dei social media che sarà insostituibile per la ricostruzione della nostra epoca e senza la quale la storia ne risulterebbe irrimediabilmente menomata. Si auspica che iniziative di questo genere vengano messe in campo anche da altre agenzie formative e dalle istituzioni di conservazione della memoria, al fine di garantire la conservazione per il futuro di queste nuove ed insostituibili fonti per la storia.

BIBLIOGRAFIA

- [1] «Archive-It». s.d. Archive-It. <https://archive-it.org>.
- [2] Bracciotti, Lorenzana. 2019. «Il Web Archiving. Conservazione e uso di una nuova fonte». *Officina della storia* (blog). 10 gennaio 2019. https://www.officinadellastoria.eu/it/2019/01/10/il-web-archiving-conservazione-e-uso-di-una-nuova-fonte/#_edn2.
- [3] Costa, Miguel, Daniel Gomes, e Silva. 2017. «The evolution of web archiving». *International Journal on Digital Libraries* 18 (settembre): 191–205.
- [4] «Cyotek WebCopy». s.d. <https://www.cyotek.com/cyotek-webcopy>.
- [5] «GitHub». s.d. GitHub. <https://github.com/webrecorder/wacz-spec>.
- [6] «GitHub - nla/htrack2warc». s.d. GitHub. <https://github.com/nla/htrack2warc>.
- [7] «HTTrack». s.d. HTTrack. <https://www.httrack.com>.
- [8] «International Internet Preservation Consortium (IIPC)». s.d. Netpreserve. <https://netpreserve.org/>.
- [9] «Internet Archive». s.d. <https://archive.org>.
- [10] «ISO 28500:2017 Information and documentation — WARC file format». 2017. ISO. <https://www.iso.org/obp/ui/#iso:std:iso:28500:ed-2:v1:en>.
- [11] «ISO/TR 14873:2013 Information and documentation — Statistics and quality issues for web archiving». 2013. ISO. <https://www.iso.org/obp/ui/#iso:std:iso:tr:14873:ed-1:v1:en>.
- [12] Landino, Costantino, e Lina Marzotti. «Perché dovremmo pensare al web archiving». *Forum PA - Cantieri PA* (blog). 20 marzo 2019. <https://www.forumpa.it/pa-digitale/gestione-documentale/perche-dovremmo-pensare-al-web-archiving>.
- [13] «List of Web archiving initiatives». s.d. Wikipedia. https://en.wikipedia.org/wiki/List_of_Web_archiving_initiatives#Archived_data.

- [14] Lorenzana, Bracciotti. 2020. «Pandemia e web archiving. Conservare le fonti online #igiornidellapandemia». *Il mondo degli archivi* (blog). 2 maggio 2020. <http://www.ilmondodegliarchivi.org/rubriche/archivi-digitali/815-pandemia-e-web-archiving-conservare-le-fonti-online-igiornidellapandemia>.
- [15] Storti, Chiara. 2019. «Web archiving, “sfida culturale”: il servizio della Biblioteca Nazionale Centrale di Firenze». *Forum PA - Cantieri PA* (blog). 12 giugno 2019.
- [16] Unesco. 2003. «Charter on the Preservation of Digital Heritage». *Unesco* (blog). 15 ottobre 2003. http://portal.unesco.org/en/ev.php-URL_ID=17721&URL_DO=DO_TOPIC&URL_SECTION=201.html.
- [17] «Web and social media archiving and preservation». s.d. <https://site.unibo.it/web-and-social-media-archiving-and-preservation/it>
- [18] «Web-Archiving: DPC Technology Watch Report». s.d. DPC online. <https://www.dpconline.org/digipres/discover-good-practice/tech-watch-reports>