

Accumulatori digitali compulsivi,

Autore **Giancarlo Buzzanca**

In pubblicazione sul numero 118 della rivista *Kermes*,
nella rubrica *Internet èer il restauro*

Poniamo il caso che qualcuno, preso da indicibili motivazioni, si trovi a consultare il sito www.abaq-informatica.info con cui veicolo il materiale che, come docente di "Informatica di base" presso l'Accademia delle Belle Arti de L'Aquila, utilizzo per le mie lezioni.

Il signor qualcuno troverà sulla *home page* la piccola sezione intitolata "Personaggio del giorno, dall'inizio della storia dell'informatica ad oggi", che contiene rapide indicazioni su:

1. Ada Byron Lovelace, matematica e programmatrice inglese
2. Alan Turing, uno dei padri dell'informatica
3. Margaret Hamilton, Direttrice del Software Engineering Division del MIT Instrumentation Laboratory,
4. Vinton Cerf, uno dei "padri di Internet" avendo prodotto la suite di protocolli TCP/IP.
5. Ted Nelson, colui che coniò il termine "ipertesto"
6. Tim Berners-Lee, l'inventore del web
7. Robert Cailliau, co-inventore del World Wide Web
8. Roy Tomlinson, l'uomo che inventò l'e-mail
9. Richard Stallman, esponente del movimento del Free Software
10. Jimmy Wales, fondatore del progetto Wikipedia
11. Brewster Kahle, fondatore di Internet Archive e della Open Content Alliance,
12. Linus Torvalds, autore della prima versione del kernel Linux Open Source

I personaggi citati appartengono tutti ad un ambito strettamente informatico, che costituisce l'alveo di appartenenza di internet. La lista dei nomi, tutti personaggi di primo piano, suscita un interesse davvero particolare dal punto di vista di costruzione della cultura informatica come costruzione della struttura logica/materiale, trasmissione dei valori e conservazione dei principi fondanti.

Aggiungo che le singole vicende umane di questi personaggi ci aiutano a delineare la storia generale e a definirne le caratteristiche.

E' per questo che butto all'interno di questa mischia la sconosciuta a quasi tutti Abigail (meglio nota come Abbie) Grotke che ricopre il ruolo di *web archiving team lead at the Library of Congress in Washington D.C.* Lo spunto è generato da un articolo¹ pubblicato da *Steven Kurutz* sul New York Times il 7 aprile 2020 e tradotto e pubblicato da *la Repubblica* il 14 aprile. L'occasione che ha generato l'intervento di Kurutz è il 20° anniversario dell'inizio del progetto dell'archiviazione delle risorse Web proprio presso la Library of Congress.

Diamo un "volto" a Abigail Grotke, oggi cinquantaduenne, che si laurea in Storia dell'arte nel 1991 ottenendo rapidamente uno stage presso il National Museum of American Art dello Smithsonian Institution per poi passare a lavorare , nel 1997, presso la *Library of Congress in Washington D.C.* con il compito di *"build the information superhighway through the Library of Congress!* Un compito davvero impegnativo che puntava alla messa a punto di un'autostrada di raccolta di informazioni in primo luogo per garantirne la sopravvivenza. Il web non è chiacchiericcio, non è gestione social, non è forma ma è la nostra attuale sostanza. Si potrebbe addirittura affermare che informazione e web (o meglio ancora internet in tutte le sue possibili evoluzioni) siano oggi sinonimi.

Negli ultimi 20 anni, una piccola squadra di archivisti presso la Biblioteca del Congresso ha raccolto il contenuto della rete silenziosamente e con rispetto. L'iniziativa è nata dal desiderio di raccogliere e preservare materiali ad accesso aperto dal web, in particolare i contenuti prodotti dal governo degli Stati Uniti durante le vicende elettorali.

Archiviare l'informazione è storia antica che ha sempre visto il bibliotecario/archivista leggere quotidiani e riviste identificando e "cerchiando" gli articoli di interesse. Il team dell'archivio web della Biblioteca del Congresso è partito con la definizione di uno staff di cinque persone, tutte donne, ed una serie di punti di riferimento esterni al team ma collocati all'interno della amministrazione del Congresso cui era ed è demandato il compito di identificare le fonti di informazioni potenzialmente interessanti.

It was in 2000 that the Library of Congress embarked on a web preservation pilot project, which eventually became the Library's web archiving program. An acronym to describe the pilot program was born to align with a beautiful mosaic in the Jefferson Building: "MINERVA: Mapping the Internet Electronic Resources Virtual Archive" From our records, the pilot program activities began around this time in 2000, and after some early test crawls, the first collection, related to the Election 2000, began in August 2000.²

Ma in generale l'approccio a internet/web è profondamente diverso da quello operato correttamente da Abigail Grotke e dal suo team all'interno della Biblioteca del Congresso. Nella generalità dei casi, infatti, il metodo di approccio si snoda nell'assoluta prevalenza dell'inconsapevolezza dell'agire che, per sua natura, è non scientifica.

Ma, *ça va sans dire*, che l'ampia diffusione dell'improvvisazione elevata a metodo non ne garantisce la qualità. E' come dire che il libro *La tua seconda vita comincia quando capisci di averne una sola* di Raphaëlle Giordano e S. Arena, è il libro più venduto nel 2019 in Italia, ma non è detto che sia il migliore tra quelli pubblicati nell'anno. E' semplicemente il più venduto.

Ma torniamo un passo indietro, qual è la consapevolezza che noi abbiamo su queste tematiche ragionando in termini istituzionali?

Per rispondere a questo quesito, mi permetto una premessa, se vogliamo ironica, rispetto al fatto che subito dopo la partenza a Washington del progetto MINERVA, appena citato, in

ambito europeo fu lanciato un progetto importante e oggettivamente pieno di risultati, in un ambito diverso, quello dell'accessibilità e fruibilità in rete dei beni culturali, da quello in cui è intervenuta la Biblioteca del Congresso, ma con nome stranamente identico, pur attingendo ad un acronimo diverso:

Il progetto MINERVA - Ministerial Network for Valorising Activities in digitisation è stato finanziato dalla Commissione europea nell'ambito del programma IST (Tecnologie per la società dell'informazione) (...) e riunisce i ministeri dei paesi europei preposti alla cultura, coordinati da quello italiano. La sua nascita è il risultato di una stretta collaborazione tra la Commissione Europea e gli Stati Membri, con l'obiettivo di facilitare la creazione di una visione comune nella definizione delle azioni e dei programmi nel campo dell'accessibilità e fruibilità in rete dei beni culturali. Il progetto ha avuto inizio nel marzo 2002; la sua prima fase è terminata con successo nel luglio 2005, ottenendo risultati ben oltre le aspettative.³

Minerva, (per i greci Ἀθηνᾶ) figlia di Giove e della sua prima moglie Metide, era la dea romana della guerra, della saggezza, della poesia e della medicina nonché protettrice degli artigiani. Se riportassimo in auge il politeismo greco sapremmo già come aggiornare la sua figura cui, evidentemente, siamo devoti per il significato veicolato, su entrambe le sponde dell'Oceano Atlantico. Chi costruisce il web, chi lavora sulla sua storia e sulle sue regole tecniche, chi lavora per garantire il mantenimento in vita delle informazioni ha di fronte a sé un mare immenso. Immenso per l'incrociarsi dei problemi tecnici, immenso per le sue dimensioni fisiche.

In Europa scimmiettiamo l'America? Sarebbe triste ammetterlo perché materiale base, competenze e inventiva non ci mancano di certo.

Ma torniamo alla *Library of Congress* che ha già accumulato oltre 2.129 petabyte (un petabyte equivale a 10^{15} byte) di dati o, in altri termini, 18 miliardi di documenti digitali. E questo è solo un frammento di Internet.

Meglio allora non citare, per non rimanere turbati, i dati di Internet Archive che sono al limite dell'incredibile per la ricchezza, complessità, e disponibilità di questo patrimonio digitale.

The Internet Archive is an American digital library with the stated mission of 'universal access to all knowledge.' It provides free public access to collections of digitized materials, including websites, software applications/games, music, movies/videos, moving images, and millions of books. In addition to its archiving function, the Archive is an activist organization, advocating a free and open Internet.

The Internet Archive allows the public to upload and download digital material to its data cluster, but the bulk of its data is collected automatically by its web crawlers, which work to preserve as much of the public web as possible. Its web archive, the Wayback Machine, contains hundreds of billions of web captures. The Archive also oversees one of the world's largest book digitization projects.

Archiviare il web è davvero una operazione quasi impossibile. Archiviare, organizzare, selezionare il patrimonio in ambito vicino alla cultura del restauro e della conservazione direi non di meno. Ma è azione che deve essere compiuta senza alcun indugio.

Progetti di conservazione del web stanno crescendo anche in Italia ma mi rimane sempre questo terrificante dubbio su chi fa che cosa e sugli strumenti adottati. "*Nella vastità del web, qual è il campionamento di cose che possiamo mettere insieme per dimostrare cosa sta succedendo ora?*" su chiede John Fenn, responsabile della ricerca e dei programmi dell'American Folk life Center poiché è uno dei circa 80 collaboratori che forniscono suggerimenti per l'archivio della biblioteca, nel caso di Fenn, per la collezione Web Cultures.

I criteri di selezione tipicamente utilizzati dagli archivisti della stampa - valore per i futuri studiosi ed unicità del materiale - si applicano ancora agli archivisti del web, sebbene l'elevatissimo tasso di estinzione della materia digitale entra nel processo decisionale in maniera ancor più preponderante che non per la carta. Ma anche questo tema lo evito, ora, tranne che rimandare i miei affezionati quattro lettori (sempre che dopo questo articolo non scendano a tre) alla quantità piuttosto robusta di articoli che ho pubblicato su questo tema.

Ma quali sono stati i temi che hanno guidato la costruzione del virtual archive del progetto *MINERVA nell'ambito della Biblioteca del Congresso?*

Il primo materiale nell'archivio risale alle elezioni del 2000, quando l'archivio web era ancora un programma pilota. Dopo gli attacchi terroristici dell'11 settembre, quando memoriali strazianti e accesi dibattiti politici si sono svolti online, la biblioteca ha riconosciuto la necessità di un record digitale ufficiale.

Per anni, la raccolta è stata improntata ai principali eventi di cronaca: la guerra in Iraq, le elezioni del 2004. Poi, intorno al 2009, è arrivato un approccio più continuo e ampliato che ha cercato di riflettere il web in tutta la sua vertiginosa novità. Bisogna dire che lavoro non manca e, ad esempio, nel tempo di COVID-19 ci sarebbe molto da archiviare.

Ho avuto, in passato, occasione di scrivere sulla storia del web e sulla sua conservazione. Prendiamo il caso del sito del MiBACT e chiediamoci dove e come sia possibile vedere e consultare la storia del medesimo sito web. La risposta è sulla Wayback Machine.

La Wayback Machine, progetto dell'Internet Archive, ha scansionato, a partire dal 1996, Internet, conservando oltre 411 miliardi di pagine Web (dati della fine del 2018, gli ultimi disponibili). Ed i siti prodotti e attivi prima del 1996? Possiamo solo immaginarli ammesso e non concesso che le persone a suo tempo impegnate sul campo abbiano avuto la consapevolezza di registrare ed archiviare le versioni dei siti.

"Non abbiamo tempo di fermarci e renderlo più facile da usare. Stiamo solo cercando di raccogliere tutto prima che scompaia. (...) I nostri archivi sono semplicemente enormi e continuano a crescere e crescere".

Oltre a dirigere il team, l'altro compito della Abbie Grotke che si definisce a capo di un team di **accumulatori digitali compulsivi**, è di sensibilizzare il pubblico sull'esistenza dell'archivio. Il sito web dell'archivio è disponibile per chiunque disponga di una connessione Internet, ma dopo 20 anni rimane sottoutilizzato dal pubblico in generale e dagli studiosi a cui può essere più utile. Ed in più l'archivio, dice la Grotke, è attualmente "leggermente impenetrabile" e questo è conseguenza delle risorse limitate e dell'enorme quantità in continua espansione dei contenuti digitali.

Torniamo alle parole, tradotte in italiano, della Abbie Grotke :

L' "archivio internet" di solito attira l'attenzione durante i party. La risposta più lunga è "sono il responsabile del team di archiviazione Web per il programma di archiviazione web della Biblioteca". L'archiviazione Web è il processo mediante il quale utilizziamo strumenti speciali per creare copie dei contenuti Web per la conservazione e l'accesso da parte di futuri ricercatori. Conserviamo una vasta gamma di contenuti pubblicati sul Web - non solo siti Web, ma anche singoli documenti, video, audio, immagini, social media, ecc. Gli archivi Web della Library of Congress sono organizzati in raccolte tematiche e basate su eventi e contengono siti Web documentare una varietà di organizzazioni statunitensi e internazionali che rappresentano una vasta gamma di argomenti e aree tematiche, come l'Archivio Web Blawgs legale in corso presso la Biblioteca della Legge, l'Archivio Web del Congresso degli Stati Uniti e l'Archivio Web dei tribunali federali. Anche se non seleziono personalmente il contenuto degli archivi della Biblioteca (...), ricopro un ruolo di gestione del programma, supervisionando le attività del nostro team di quattro membri del personale dei Servizi della Biblioteca. Ci assicuriamo che il contenuto identificato dai responsabili delle raccomandazioni per le raccolte sia archiviato (nel miglior modo possibile con gli strumenti che abbiamo), archiviato per la conservazione a lungo termine e reso disponibile per l'uso nella ricerca. In questo ruolo, posso anche lavorare con un certo numero di colleghi di archiviazione web negli Stati Uniti e nel mondo attraverso gruppi come l'Internet Internet Preservation Consortium e il progetto collaborativo End of Term Web Archive. Ogni giorno è un po' diverso con l'archiviazione web, il che lo rende eccitante.⁴

Ed io, molto banalmente, commento ... ma quando avremo anche noi, in Italia, una struttura simile?

E se poi volessimo pensare all'ambito più specifico della conservazione e del restauro, e a quanta necessità ci sarebbe di archiviare, organizzare, selezionare il patrimonio di informazioni presenti sul web divenendo anche noi *accumulatori digitali compulsivi*? Mi verrebbe da rispondere "ai posteri l'ardua sentenza" se non fosse che i posteri in questione siamo noi.

Sintesi

*Considerate la vostra semenza:
fatti non foste a viver come bruti,
ma per seguir virtute e canoscenza
(Dante)*

Le immagini

Immagine 1 The web archiving team, presso la Library of Congress .Crediti...Justin T. Gellerson per The New York Times

Immagine 2 – Il server che ospita il Web archive. Crediti...Justin T. Gellerson per The New York Times

Immagine 3 – La home page del sito MINERVA : (Mapping the INternet Electronic Resources Virtual Archive) nel 2001

Immagine 4 – La crescita del Web archive della Library of Congress negli ultimi 20 anni

¹¹<https://www.nytimes.com/2020/04/07/style/internet-archive-library-congress.html>

²<https://blogs.loc.gov/thesignal/2020/04/happy-birthday-to-lcwa-celebrating-the-20th-anniversary-of-web-archiving-at-the-library-of-congress/>

³https://www.beniculturali.it/mibac/export/MiBAC/sito-MiBAC/Contenuti/Progetti/Archivio/Progetti-Europei/visualizza_asset.html_563517191.html

⁴<https://blogs.loc.gov/law/2019/01/an-interview-with-abigail-grotke-web-archiving-team-lead/>





MINERVA

Mapping the INternet Electronic Resources Virtual Archive

Browse the Archives *(available onsite only)*

- [By Title](#)
- [By Subject](#)
- [By URL](#)

Find out More About the Project

- [Project Statement](#)
- Selection Criteria
- Cataloging /Description
- Technical Architecture
- [Copyright Issues](#)
- Partners
- What's New



*A Web Preservation Project
of the Library of Congress*

*Mosaic of Minerva, Great Hall
of the Library of Congress*



