

Federico Meschini

RETI, MEMORIA E NARRAZIONE

ARCHIVI E BIBLIOTECHE DIGITALI TRA
RICOSTRUZIONE E RACCONTO



Prima edizione: luglio 2018

ISBN: 978-88-7853-806-1

ISBN ebook: 978-88-7853-659-3

IN COPERTINA: incisione della città di Königsberg e dei suoi sette ponti del 1581.

Riproduzione vietata ai sensi di legge
(art. 171 della legge 22 aprile 1941, n. 633)

Impaginazione e grafica:
Fabiana Ceccariglia per Tramaglio
www.tramaglio.it

Edizioni **SETTE CITTÀ**

Via Mazzini 87
01100 Viterbo
tel +39 0761 304967
fax +39 0761 1760202

info@settecitta.eu
www.settecitta.eu

SOMMARIO

p. 9	Introduzione
17	1. Documenti, dati e metadati. Il WordWideWeb come spazio conoscitivo
93	2. Ricostruire e raccontare la memoria: il caso della Grande Guerra
145	3. Ricostruire e raccontare il viaggio: la letteratura odeporica
209	Ringraziamenti
213	Bibliografia

INTRODUZIONE

La pervasività degli archivi e delle biblioteche digitali ha ormai da tempo superato le strette cerchie degli specialisti accademici o, più in generale, degli esperti di settore, raggiungendo un pubblico ampio, sebbene stratificato, frammentato e non certo classificabile con l'etichetta "generalista", in quanto sempre più interessato ad argomenti specifici, seppure con le dovute diversità tra i vari livelli di approfondimento. Questo certamente grazie alla loro crescita sia orizzontale, ad un livello quantitativo, con una copertura progressiva dei vari campi dello scibile, sia verticale, la dimensione qualitativa di ogni singolo sistema informativo classificabile come archivio o biblioteca digitale, ma soprattutto a causa delle capacità automatiche di disseminazione e propagazione capillare su piattaforme di fruizione eterogenee, tipiche ormai di un certo modo di intendere l'informazione digitale in rete.

Un aspetto che verrà esaminato in dettaglio è il superamento degli aspetti relativi all'autoreferenzialità dei dati e alla loro conservazione in contenitori ermetici, così come successo in generale per i sistemi informativi di cui archivi e biblioteche digitali sono un'evidente specializzazione. Ciò è avvenuto da un lato grazie ai paradigmi del Web 2.0, in particolare per quello che riguarda l'aspetto sintattico e dei protocolli d'interoperabilità, e dall'altro per mezzo del Semantic Web, per l'aspetto seman-

tico e l'utilizzo di particolari vocabolari di metadati relativi ai domini della conoscenza. Quest'argomento verrà trattato senza scendere eccessivamente in dettagli tecnici, che esulerebbero dallo scopo di questa pubblicazione, ma dando al tempo stesso alla tecnologia il giusto ruolo in quella che è una continua evoluzione degli ambienti digitali di rete.

Questo fenomeno va inoltre contestualizzato in uno scenario d'immersione e fruizione informativa che, in particolare per le nuove generazioni, ha sempre più luogo in questi stessi ambienti; naturalmente non senza conseguenze, soprattutto per quel che riguarda le modalità di organizzazione dell'informazione e di conseguenza della conoscenza, le cui condizioni apparentemente fisiologiche sembrano corrispondere sempre di più ai criteri combinati di alto tasso di frammentazione ed elevata velocità di fruizione, e, verrebbe da pensare come logica conseguenza, di scarsa profondità e debole, se non del tutto assente, struttura; caratteristiche, queste ultime, sulla cui effettiva contingenza è necessario interrogarsi, in quanto, come vedremo, strettamente collegate all'oggetto di studio in questione¹.

Sia a livello di contenuto sia di espressione, sia per ciò che riguarda i singoli oggetti contenuti sia l'ambiente sovrastante (che dà forma alla natura stessa di questi sistemi), archivi e biblioteche digitali sono perciò una risposta a quel bisogno di complessità sottolineato da Gino Roncaglia nel parlare dell'era delle cattedrali², e che perciò va in controtendenza rispetto allo scenario appena descritto.

¹ Per una prima introduzione al tema della complessità relativamente alle forme di organizzazione dell'informazione digitale cfr. G. Roncaglia, *Social network e riconquista della complessità: il ruolo delle biblioteche* in "Biblioteche oggi", XXXII, 2014, 5, pp. 4-8, <<http://www.bibliotecheoggi.it/2014/20140500501.pdf>> e successivamente in forma più articolata G. Roncaglia, *L'età della frammentazione. Cultura del libro e scuola digitale*, Roma-Bari, Laterza, 2018.

² *Ivi*, pp. 28-34.

In un'ottica evolutivo-diacronica il paradigma di alto livello attualmente in auge di questi sistemi conoscitivi si afferma a partire dai primi anni '90, avendo quindi come infrastruttura di base il WorldWideWeb, tralasciando perciò i paradigmi precedenti, molti dei quali si sempre basati, a basso livello, sull'infrastruttura di Internet, come la prima versione del Progetto Gutenberg, realizzata negli anni '70 e successivamente trasmigrata sul Web, ma accessibili tramite protocolli di livello applicativo (secondo il modello ISO/OSI), tra cui il telnet, caratterizzati da un livello di usabilità decisamente molto più basso e un'ardua curva di apprendimento, soprattutto considerando lo sviluppo nel settore del design delle interfacce utente.

Si è specificato di alto livello per due motivi; il primo concerne l'evoluzione continua del web, il suo costante aggiornamento sia a livello di tecnologie sottostanti sia, come conseguenza, di relative funzionalità, generando quindi diverse configurazioni di questo paradigma generale, sia sull'asse diacronico sia su quello sincronico, con nel primo caso continui aggiornamenti tecnologici che ne modificano, in maniera più o meno radicale, l'assetto e le funzioni e nel secondo caso, più raro ma non per questo trascurabile, la presenza di implementazioni eterogenee degli stessi contenuti.

Circoscrivendo i numerosi possibili esempi nell'ambito delle *Digital Humanities*, di cui ci occuperemo con una certa consistenza, al primo caso, appartengono il *Rossetti Archive*³ o il *Blake Archive*⁴, incentrati rispettivamente, così come il nome lascia intuire, sulle opere di Dante Gabriel Rossetti e di William Blake. Entrambi erano basati nelle loro prime versioni su DynaText, un sistema proprietario per la pubblicazione di testi codificati in SGML, e una volta convertiti i contenuti in XML sono passati all'utilizzo di strumenti open source, come il motore di ricerca Lucene o il database nativo XML eXist. Sincronicamen-

³ *The Rossetti Archive*, <<http://www.rossettiarchive.org/>>.

⁴ *The William Blake Archive*, <<http://www.blakearchive.org/>>.

te è possibile avere diverse applicazioni con gli stessi contenuti sottostanti, e basate su piattaforme specifiche che renderanno disponibili le varie funzionalità in maniera idiosincronica, ponendo giustamente la questione se sia legittimo considerarli o meno come lo stesso sistema informativo o piuttosto come declinazioni eterogenee di uno stesso contenuto o di uno o più insiemi di contenuti. Come ad esempio la biblioteca digitale *Perseus*⁵, incentrata sui testi classici, i cui contenuti sono stati utilizzati con *Philologic*, uno strumento di ricerca e analisi testuale⁶, creando così *Perseus under Philologic*⁷ con, rispetto alla sua versione tradizionale, scarse se non nulle funzionalità di visualizzazione e fruizione dei testi ma dotato di funzioni avanzate di ricerca; oppure sconfinando nel settore collegato delle edizioni digitali, considerando anche il relativo cambio di supporto, il *Wittgenstein Nachlass*⁸, edizione critica digitale su DVD dell'intero lascito documentale di Ludwig Wittgenstein, è stato reso parzialmente disponibile sul web tramite il progetto *WittgensteinSource*⁹. E naturalmente non mancano le combinazioni dell'aspetto diacronico con quello sincronico, in quanto sia il *Perseus Project* sia *WittgensteinSource* sono stati più volte tecnologicamente aggiornati se non del tutto rinnovati, così come nel noto paradosso della nave di Teseo.

Quest'ultima considerazione ci porta al secondo motivo, riguardante la natura stessa degli archivi e delle biblioteche digitali, la realizzazione concreta di una singola istanza appartenente a questo insieme astratto: una complessa macchina tecno-

⁵ *Perseus Digital Library*, <<http://www.perseus.tufts.edu/hopper/>>.

⁶ *PhiloLogic*, <<http://www.lib.uchicago.edu/efts/ARTFL/philologic/>>.

⁷ *Perseus under PhiloLogic*, <<http://perseus.uchicago.edu/>>.

⁸ *Wittgenstein's Nachlass. The Bergen Electronic Edition (BEE)*, <http://wab.uib.no/wab_BEE.page>.

⁹ *WittgensteinSource*, <<http://www.wittgensteinsource.org/>>.

logica che, realizzata con i vari componenti hardware e software al momento disponibili, implementi quelle particolari funzioni che permettano di connotare e denotare lo specifico sistema informativo con queste affascinanti quanto evocative etichette¹⁰.

Un'ulteriore caratteristica di questi sistemi è l'attenzione verso l'architettura sottostante, il cui *telos* è dare stabilità e contesto perdurante a quegli oggetti la cui natura digitale sembra conferire loro una durabilità fortemente transeunte, condizione di cui soffre anche la maggior parte delle piattaforme di pubblicazione sul web, incentrate principalmente sulla velocità di disseminazione piuttosto che sulla longevità della conservazione.

Questa dicotomia apparentemente divergente tra Content Management Systems (CMS) generici, come Drupal o Wordpress, motori sottostanti la maggior parte dei siti web, e piattaforme specializzate, come Fedora o Dspace, molto sentita nei primi anni duemila, si è evoluta verso un'integrazione collaborativa tra questi due mondi; va ricordato infatti come una delle caratteristiche principali dell'informatica sia proprio la capacità di poter ridefinire continuamente i propri scenari d'utilizzo, seppur sempre all'interno di ben determinati confini logici. Esempio rappresentativo di tale integrazione in cui i CMS hanno il ruolo di interfaccia finale verso gli utenti e di gestione dei meccanismi di pubblicazione di ciò che è stato immesso in appositi repository è Islandora¹¹, molto diffuso in Nord America e combinazione dei già citati Drupal e Fedora; entrambi

¹⁰ Anche in questo caso, una discussione sulla natura di questi sistemi è un qualcosa di altro rispetto agli scopi di questo volume: non è possibile però non citare il seminale contributo di Payette e Lagoze in cui ne vengono descritti i componenti chiave architettureali, cfr. S. Payette, C. Lagoze, *Flexible and Extensible Digital Object and Repository Architecture (FEDORA)* in *Lecture Notes In Computer Science: Proceedings of the Second European Conference on Research and Advanced Technology for Digital Libraries*, 1513, 1998, pp. 41-59, <<http://arxiv.org/abs/1312.1258/>>.

¹¹ *Islandora*, <<http://islandora.ca/>>.

questi strumenti sono sì eterogenei per finalità e tecnologie ma caratterizzati da architetture aperte ed estendibili, un assetto strutturale che, seppure non privo di complessità, ne favorisce l'interazione¹².

Un tratto distintivo impossibile da non notare è quello di una duplice natura contenutistica, che riproduce simbolicamente la contrapposizione tra scienze esatte e scienze umane che la rivoluzione gutenberghiana ha contribuito ad acuire: da un lato la comunicazione scientifica e dall'altro la disseminazione e preservazione del patrimonio culturale. Questa duplicità è ben rappresentata da *arXiv*¹³, l'archivio ad accesso aperto diffusosi inizialmente nella fisica e successivamente esteso ad altre discipline, tra cui matematica e informatica, e dal già menzionato *Rossetti Archive*, entrambi progettati, realizzati e continuamente aggiornati, a partire dai primi anni novanta. Nel primo caso l'oggetto informativo principale, corredato da metadati e attorno al quale ruota tutta l'architettura sovrastante, è rappresentato da un documento PDF, il formato elaborato da Adobe nei primi anni '90 per rappresentare fedelmente l'assetto tipografico di un documento. Il PDF è a sua volta un sottoinsieme del PostScript, il linguaggio di programmazione alla base della rivoluzione del Desktop Publishing della metà degli anni '80, e con il quale viene realizzata, a parte rarissime eccezioni, la totalità dell'attuale produzione editoriale cartacea. Nel secondo caso, al contrario, data sia la diversa tecnologia sottostante il supporto originario sia il fatto di avere a che fare, la maggior parte delle volte, con un corpus testuale più ristretto e ben definito, come le opere di un particolare autore, si preferisce un approccio in cui il testo viene rappresentato tramite linguaggi descrittivi di marcatura, in particolare XML. In questo

¹² A conferma di ciò, la stessa combinazione si può ritrovare altrove, tra cui il sistema sottostante la *Sapienza Digital Library* <<http://sapienzadigitallibrary.uniroma1.it/>>.

¹³ *arXiv*, <<http://arxiv.org/>>.

modo è possibile definire una struttura testuale a diversi livelli di granularità e creare relazioni con altri blocchi informativi, eventualmente appartenenti anche ad altri codici comunicativi, realizzando così complessi grafi conoscitivi. Questa pratica, considerata la norma nella digitalizzazione del patrimonio culturale, a causa dell'inevitabile processo di ri-mediazione e delle buone pratiche sviluppate negli anni, sta prendendo piede anche nella produzione *ex novo* di contenuti scientifici, sia a livello di monografie nelle *Humanities* sia di singoli articoli nelle *Hard Sciences*, grazie anche, soprattutto nel primo caso, alle piattaforme e alle fonti primarie rese disponibili, queste ultime proprio tramite gli archivi e le biblioteche digitali¹⁴. Questo passaggio dal paradigma tipografico a quello computazionale non può non comportare, seppure con tempistiche e modalità variabili, un avvicinamento tra scienze esatte e scienze umane, a causa dei principi logico-matematici alla base della natura di questa nuova tipologia di pubblicazioni, la cui funzione strategica va quindi considerata a livello di processo e non solo di prodotto.

Infine, qualche informazione sulla struttura di questo volume, in quanto caratterizzata da un fenomeno abbastanza diffuso nella codifica testuale: la presenza di gerarchie multiple appartenenti a livelli diversi. La prima è quella logica, divisa in tre capitoli principali, ognuno fruibile autonomamente. Nel primo viene introdotto il WorldWideWeb nel suo ruolo di ambiente conoscitivo, i processi evolutivi e le innovazioni tecnologiche che hanno portato questo sistema ipertestuale locale e sperimentale a una diffusione globale e capillare, andando a inglobare e influenzare tutte le altre modalità comunicative, mantenendo una costante attenzione al ruolo delle risorse digitali culturali e alle loro modalità di disseminazione.

¹⁴ *Scalar*, <<http://scalar.me/anvc/scalar/>>, è un interessante strumento per la creazione di pubblicazioni arricchite da elementi multimediali, con una base di utenti, e conseguenti utilizzi concreti, in costante crescita, vedi *A selection of projects authored with Scalar*, <<http://scalar.me/anvc/scalar/showcase/>>.

Questo primo capitolo rappresenta la prima parte del livello concettuale, l'altra gerarchia, mentre della seconda ne fanno parte i rimanenti due capitoli, in cui una volta stabilite le premesse necessarie sul ruolo e le capacità del Web come strumento di ricostruzione e di racconto, si passa ad analizzarne le applicazioni al patrimonio culturale. Per dare una certa consistenza ed evitare il senso di frammentazione, data l'evidente estensione dell'argomento, sono state effettuate due scelte, in relazione tra di loro. La prima è sull'utilizzo di due principi, quello temporale e spaziale, ognuno rappresentato da un capitolo, e la seconda su di una loro declinazione in casi concreti sufficientemente rappresentativi, rispettivamente la prima guerra mondiale e la letteratura odepórica. Ciò che rimane costante è l'assetto metodologico, basato sull'analisi di diversi progetti e pratiche, cercando di individuare i vari fattori comuni e tracciare così un cerchio partendo da un insieme di punti.

Le riflessioni contenute nella pagine successive si basano su quasi vent'anni di esperienze, letture e riflessioni, e rappresentano quindi uno spaccato non da poco, seppure settoriale, sul WorldWideWeb. La conclusione più rilevante è come, non solo nei prossimi vent'anni bensì in tutti quelli in cui questo medium continuerà ad essere utilizzato e sviluppato, il recupero e la trasmissione del passato saranno un qualcosa di fondamentale: le conseguenti ricadute positive di queste attività non rimangono circoscritte ad un singolo progetto ma si propagano a livello sistemico, rendendo oltremodo necessari considerazioni e valutazioni critiche sull'uso della tecnologia digitale, tanto più inevitabili quanto più quest'ultima si diffonde capillarmente nella trasmissione e conservazione della conoscenza¹⁵.

¹⁵ Ultima consultazione dei siti web 30 giugno 2018.