

Archivi, tecnologie semantiche e Intelligenza Artificiale. Conoscenza partecipata al servizio del territorio

Giorgia Di Marcantonio*

Introduzione

Dagli anni Novanta ad oggi, il legame tra archivi e Tecnologie dell'Informazione e della Comunicazione (ICT) ha subito una metamorfosi profonda, coinvolgendo non solo i processi con cui gli archivi vengono prodotti ma anche come essi sono gestiti, conservati e restituiti. La diffusione delle ICT ha avuto un impatto considerevole sia sugli archivi digitalizzati che su quelli digitali, coinvolgendo le professionalità che ruotano attorno alle dinamiche documentarie e arricchendo la disciplina di nuovi strati di complessità¹.

Limitatamente agli archivi storici, i processi di costruzione delle descrizioni e delle relazioni archivistiche, grazie a software specifici e sistemi informativi

* Università degli Studi di Napoli, L'Orientale.

¹ G. Bonfiglio Dosio, *La professione dell'archivista nell'era digitale*, in *Archiviare il futuro: riflessione e ricerche per una materia in divenire*, a cura di G. Volpato, F. Formiga, Università degli Studi di Verona, Verona, 2008, pp. 59-64; G. Di Marcantonio, *L'archivista: una figura in costante trasformazione*, «Officina della storia», 19, 2019, <<https://hdl.handle.net/11393/298276>>.

dedicati², hanno favorito la disseminazione di dati e informazioni che in tempi remoti potevano essere trasmessi solo tramite strumenti analogici. Da un certo punto di vista, l'evolversi del rapporto tra archivi e tecnologie ha in qualche modo agevolato l'accesso alle risorse, rendendo i complessi documentari disponibili ad un pubblico più ampio, a patto che i dati inventariali (o quelli documentali) abbiano subito un processo di riversamento e restituzione in portali raggiungibili online. Su questa *conditio sine qua non* si sono già espressi in molti³, ma è bene ribadire che non tutti gli utenti sono consapevoli che il mondo non termina laddove finisce lo schermo del loro pc e che sono tanti, anzi tantissimi, gli archivi non ancora riordinati o che non dispongono di strumenti di ricerca consultabili online.

Tuttavia, è inutile negare quanto le tecnologie abbiano ampliato il ventaglio di opportunità significative per la diffusione e la disseminazione della conoscenza, ma è altrettanto vero che sono degli strumenti le cui esternalità, positive o negative, risiederanno sempre nelle mani di chi li progetta e li utilizza. Nei meandri di algoritmi ogni giorno più complessi e di sistemi in grado quasi di replicare i processi cognitivi naturali, il fattore umano diventa sempre più determinante.

Sotto il profilo del patrimonio culturale un approccio superficiale alle ICT può condurre a delle distonie difficilmente recuperabili. Da un lato l'obsolescenza tecnologica è sempre in agguato e, se i processi documentali – digitali o digitalizzati – non sono ben progettati, si rischia di non poter più avere accesso alle informazioni e di perdere pezzi importanti della memoria individuale e collettiva di una comunità. Dall'altro la riconcorsa a creare nuovi sistemi e a nutrirli di dati può produrre una bulimia informativa che disorienta gli utenti invece che guidarli e arricchirli di nuova conoscenza.

Sostenibilità e consapevolezza dovrebbero essere alla base di ogni progetto culturale, in particolar modo se ci si orienta verso il *World Wide Web*. Tale approccio si rende ancor più necessario quando si tratta di costruire una cono-

² Per quanto riguarda i software si vedano ad esempio: Archimista, XDams, Sinapsi, etc. Per ciò che concerne i sistemi informativi si fa riferimento al Sistema Archivistico Nazionale (SAN) e tutti i vari sistemi aderenti e partner del progetto.

³ Ad esempio: C. Damiani, *La memoria rappresentata: dalla descrizione inventariale agli archivi narranti*, «Officina della storia», 19, 2019, <<https://hdl.handle.net/11591/489105>>; P. Feliciati, *Dall'inventario alla descrizione degli archivi in ambiente digitale: si possono offrire agli utenti risorse efficaci?*, relazione presenta al convegno «E-laborare il sapere nell'era digitale. Strumenti e tecniche per la gestione, la conservazione e la valorizzazione del patrimonio culturale in ambiente digitale, 22-23 novembre 2007, Montevarchi (AR)», 2008, preprint: <http://eprints.rclis.org/11159/1/Feliciati_Montevarchi1107.pdf>; L. Giuva, *Alcune osservazioni su utenti e reti archivistiche nell'era digitale*, «Archivi», 4, 2, 2009, pp. 7-20; F. Valacchi, *Problematiche descrittive e linee operative per la descrizione degli archivi in un progetto di rete*, «Archivi & Computer», 6, 3, 2007, pp. 38-49; I. Zanni Rosiello, *Su Claudio Pavone che ha preso molto sul serio il lavoro archivistico*, «Parole Chiave», 1, 2, 2019, pp. 19-34.

scenza partecipata che potrebbe essere un volano per le dinamiche sociali ed economiche dei territori. Ma su questo avremo modo di tornare.

Chaospedia *digitale*

Per avere un esempio concreto di quanto un approccio alla tecnologia precipitoso e poco ponderato possa avere delle conseguenze considerevoli, è sufficiente richiamare alla memoria quanto accaduto durante l'emergenza pandemica. Il ricorso al "digitale" per la didattica nelle scuole, per la gestione dei contagi nella sanità e per continuare a svolgere quasi ogni tipo di attività lavorativa ha avuto delle importanti conseguenze. Se da un lato alcune delle criticità possono essere attribuite alle circostanze eccezionali dell'emergenza, dall'altro lato ciò ha evidenziato una fondamentale immaturità in alcuni approcci alla digitalizzazione e alla dematerializzazione⁴. Progetti come ITsART⁵ hanno consumato ingenti fondi pubblici senza portare a risultati concreti⁶ e, successivamente, le istanze di una digitalizzazione estrema dei beni culturali nel PNRR sembrano aver guardato alla tecnologia come a una soluzione strutturale e non come a un semplice strumento⁷. Una "virtualizzazione" massiva dei beni non è sostenibile nel lungo periodo, né in termini di risorse umane né sotto il profilo economico, e denota un'ormai cronica tendenza italiana a creare tanti ambienti digitali disorganizzati, alle volte con le stesse informazioni. Tali approcci sottolineano la necessità di un indirizzo più riflessivo e strutturato nell'adozione delle tecnologie, soprattutto in un settore come quello culturale. È fondamentale sottolineare che, sebbene le ICT offrano potenzialità considerevoli per l'innovazione e l'accessibilità al patrimonio, il loro utilizzo deve essere guidato da una profonda comprensione delle esigenze specifiche del settore. È consigliabile evitare di cadere nel fascino di soluzioni solo all'apparenza ingegnose, ma che portano di fatto ad una sorta di "chaospedia" digitale che stenta ad incontrare la soddisfazione degli utenti e a valorizzare il lavoro delle professionalità coinvolte.

⁴ G. Di Marcantonio, *Le semplificazioni tecnologiche. Il rischio dei processi immaturi*, in *Dimensioni archivistiche. Una piramide rovesciata*, a cura di F. Valacchi, L. Pezzica, Editrice Bibliografica, Milano 2021, pp. 125-136.

⁵ La piattaforma "ITsART" (<<https://www.itsart.tv/it/>>, 30 ottobre 2024) era stata accompagnata da diversi comunicati. Si veda, ad esempio, quello pubblicato il 31 maggio 2021: <<https://cultura.gov.it/comunicato/20702>>, 30 ottobre 2024.

⁶ *ITsART chiude*, «Il Post», 9 maggio 2023, <<https://www.ilpost.it/2023/01/09/liquidazione-itsart-netflix-cultura/>>, 30 ottobre 2024.

⁷ G. Di Marcantonio, *Le semplificazioni tecnologiche. Il rischio dei processi immaturi*, cit.; G. Di Marcantonio, F. Valacchi, *Tra dematerializzazione e next generation: i sistemi documentari nel PNRR*, «Rivista elettronica di diritto, economia, management», 2, 2022, pp. 16-32.

Tecnologie semantiche

Per riportare ordine nel caos, le tecnologie semantiche rappresentano un ambito di ricerca interdisciplinare interessante, che si focalizza sull'applicazione di metodi e strumenti computazionali al fine di strutturare, organizzare, rappresentare e analizzare dati di qualsiasi genere⁸.

Per massimizzare l'utilizzo e l'impatto di queste tecnologie è necessario impiegare dei modelli ontologici e strutturare le informazioni in modo tale che si formalizzino delle relazioni semantiche tra entità diverse, come ad esempio persone, luoghi o eventi⁹. Grazie poi ad algoritmi di Elaborazione del Linguaggio Naturale (NLP) è possibile estrarre dati da documenti testuali e convertire queste informazioni non strutturate in rappresentazioni semantiche comprensibili non solo per l'uomo, ma in particolar modo per le macchine.

Tale modellazione *data-driven* favorisce una navigazione e un'interrogazione più intuitive ed in parte anche più efficienti. Il processo di restituzione dei dati ruota attorno a *query* semantiche-ontologiche che migliorano l'identificazione di relazioni complesse, arricchendo così l'analisi e l'interpretazione dei dati, sia lato macchina che lato utente.

Sotto il profilo archivistico queste tecnologie sembrano aprire nuove prospettive di indagine, in quanto permetterebbero di rappresentare, organizzare e ricercare i dati – al livello sia delle descrizioni che dei contenuti – in una forma più intuitiva e dinamica. Tale interesse disciplinare è alimentato dall'attenzione dell'*International Council on Archives* che, grazie al lavoro di una specifica commissione¹⁰, sta investendo tempo ed energie al fine di sviluppare un modello ontologico che sia in grado di rappresentare i fenomeni archivistici in una logica Entità-Relazioni¹¹.

Genericamente le ricadute positive dell'uso delle tecnologie semantiche per la rappresentazione di dati archivistici si tradurrebbero in una maggiore interconnettività tra bacini di informazioni, oggi isolati in singoli sistemi informativi o portali, migliorando l'interrogazione delle risorse e favorendo analisi comparative e interdisciplinari. A ciò si aggiungerebbe una maggiore intelligibilità del patrimonio informativo archivistico che verrebbe restituito non più in una forma strutturata ma in modalità aggregata e semanticamente più significativa (resa cioè anche in un linguaggio naturale).

⁸ Cfr. F. Tomasi, *Organizzare la conoscenza: digital humanities e web semantico*, Editrice Bibliografica, Milano 2022.

⁹ S. Vassallo, *From Typewriter to Bit: How Finding Aids Evolve*, «JLIS.it», 14, 3, 2023, pp. 83-104.

¹⁰ International Council on Archives, *EGAD Steering Committee*, 2016, <<https://www.ica.org/ica-network/expert-groups/egad/>>, 30 ottobre 2024.

¹¹ International Council on Archives. Records in Contexts Ontology (ICA RIC-O) version 1.0.1, <https://www.ica.org/standards/RiC/RiC-O_1-0-1.html>, 30 ottobre 2024.

Se all'apparenza tali operazioni possono sembrare lineari e di semplice applicazione, in realtà portano con sé delle profonde complessità e implicazioni sia di carattere tecnico che metodologico. Creare un sistema semantico di informazioni richiede competenze specifiche in ambito informatico e i costi di sviluppo e manutenzione di questi sistemi avanzati non sono indifferenti. A ciò si collega la difficile sostenibilità nel tempo di questi progetti, sia sul fronte dell'obsolescenza tecnologica che su quello dell'interoperabilità¹². Sul primo punto è ormai evidente quanto le tecnologie cambino rapidamente e ciò potrebbe accelerare velocemente l'obsolescenza dei sistemi. Si aggiunga a questo l'esigenza di programmare degli aggiornamenti costanti e verificare periodicamente che le risorse siano correttamente restituite e visibili agli utenti. Inoltre, l'esigenza di frammentare in atomi indipendenti l'informazione dei complessi archivistici¹³ potrebbe spingere i gruppi di lavoro a creare specifiche ontologie che limiterebbero l'aggregazione di dati in altri portali su scala più ampia.

Se da un lato, quindi, le tecnologie semantiche offrono dei potenziali benefici per la rappresentazione e la restituzione dei dati archivistici, dall'altro i rischi di tali operazioni necessiterebbero di un'attenta considerazione. Per limitare gli svantaggi associati all'uso di questi linguaggi bisognerebbe investire, tra le altre attività che andrebbero previste, in una formazione interdisciplinare costante, favorendo l'uso di piattaforme open source, standard e APIs condivisi. Attuando una combinazione di strategie mirate alla federazione delle risorse, piuttosto che alla ricerca del sistema più innovativo sul mercato, si potrebbero massimizzare i benefici dell'implementazione di questi strumenti ed eventualmente allargarli a nuovi domini.

Nuovi “giochi”, stessi problemi

Oltre alle tecnologie semantiche, le applicazioni basate sull'Intelligenza Artificiale (AI) stanno guadagnando un posto considerevole nel dibattito scientifico interdisciplinare. Questa crescente attenzione è alimentata dalle notevoli potenzialità e dall'ampio spettro di applicazioni dell'AI, che si estende a settori diversificati come medicina o finanza, fino ad arrivare all'ampio universo dell'entertainment. Le discipline umanistiche non sono rimaste immuni all'attrattiva di questi avanzati algoritmi. In ambito accademico, sia sul fronte della didattica che della ricerca, sono numerose le iniziative che puntano ad integrare l'Intelligenza Artificiale nelle loro metodologie¹⁴.

¹² *ReLoad. Repository for Linked Open Archival Data*, 2016, <<https://labs.regesta.com/progettoReload/en/>>, 30 ottobre 2024.

¹³ Come detto sopra: persone, luoghi, eventi.

¹⁴ Si veda ad esempio il progetto *InterPARES Trust AI*, 2021-2026, <<https://interparestrustai.org>>, 30 ottobre 2024.

Tale successo è stato anche stimolato dalla possibilità di interagire gratuitamente e senza avere delle specifiche competenze di dominio, con *Large Language Models* di considerevoli dimensioni come ChatGpt¹⁵. Queste sperimentazioni *homemade* hanno da subito fatto emergere gli evidenti limiti di tali strumenti che non di rado, pur di rispondere alle domande degli utenti, riportano informazioni, notizie e contenuti inaffidabili, ossia costruiti tramite un raffazzonato assemblaggio di informazioni. Studi abbastanza recenti provano a migliorare la qualità delle risposte anche grazie all'integrazione di *Retrieval Augmented Generation* (RAG) i cui esiti si stanno dimostrando molto interessanti¹⁶.

Le aree di intervento dell'AI nel settore dei beni culturali sono diverse. Senza alcuna pretesa di esaustività, si considerino ad esempio:

- La digitalizzazione avanzata. L'integrazione di scanner avanzati con software potenziati da algoritmi AI che permettono di convertire oggetti fisici in dettagliati modelli tridimensionali. Questa innovazione non si limita solo alla mera replica digitale, in quanto l'AI consente anche una classificazione semi-automatica di tali risorse, rendendo potenzialmente più agevole e intuitiva la ricerca da parte degli utenti.
- Restauro digitale. L'Intelligenza Artificiale può andare oltre la semplice analisi visuale delle opere d'arte. Algoritmi sofisticati sono in grado di rilevare segni di degrado e, sfruttando dati storici, procedere ad un restauro digitale. Questo approccio non solo conserva le opere in una forma più vicina al loro stato originale, ma le rende anche più comprensibili e fruibili dal grande pubblico.
- Valorizzazione e trasmissione del patrimonio. In questo caso le applicazioni sono diverse, ma in generale l'AI può essere d'ausilio alla disseminazione di contenuti culturali verso persone con disabilità diverse, tramite la creazione di strumenti avanzati che facilitano l'intelligibilità dei beni.

Quelli riportati sono solo alcuni esempi delle potenziali applicazioni dell'AI nel dominio culturale. Dovrebbero però essere utili a far comprendere l'esigenza di condurre un'analisi interdisciplinare su questo settore in costante evoluzione. Ciò è vero anche per il mondo accademico, nel quale è già abbastanza diffusa la tendenza a richiedere la presenza di tecnologie di frontiera nei bandi competitivi destinati al finanziamento della ricerca. Tale dinamica, anche se ancora contenuta rispetto alle dimensioni del fenomeno nel suo complesso, non è da sottovalutare per diverse ragioni. Le discipline umanistiche soffrono da molti anni di una drastica riduzione di finanzia-

¹⁵ ChatGPT, <<https://chatgpt.com>>, 30 ottobre 2024.

¹⁶ G. Di Marcantonio, *Artificial Intelligence, Large Language Models (LLMs), and Retrieval-Augmented Generation (RAG). New Tools for Accessing Archival and Bibliographic Resources*, «Bibliothecae.it», 13, 1, 2024, pp. 146-173.

menti, che spesso sono andati a favorire l'innovazione tecnologica mettendo in ombra la ricerca di base. Se sono comprensibili i tentativi di sopravvivenza del settore, è altrettanto auspicabile una riflessione sull'impiego massivo, ma soprattutto poco ragionato, di queste tecnologie che potrebbero portare ad una deriva difficilmente recuperabile a posteriori. In ambito strettamente archivistico il rischio di una sovra-interpretazione dei dati tale da condurre a conclusioni errate, fuorvianti o addirittura ad archivi del tutto inventati¹⁷ è più che plausibile e a tratti preoccupante. Con ciò non si intende porre limiti ad un progresso inevitabile e già in corso, ma semplicemente richiamare all'attenzione di professionisti ed utenti che qualsiasi fenomeno digitale dovrebbe essere guidato da un approccio consapevole e ponderato, che tenga conto in egual modo sia dei vantaggi che dei rischi cui tali operazioni possono portare.

Conoscenza partecipata al servizio del territorio

Le riflessioni fin qui riportate appaiono ancor più significative nei casi in cui le tecnologie supportino lo sviluppo di processi di conoscenza partecipata volti a favorire un'interpretazione più dinamica e soprattutto inclusiva del patrimonio culturale.

I beni culturali, come è ormai assodato, non sono oggetti passivi deputati prevalentemente allo studio e alla conservazione, ma elementi vivi che partecipano alla costruzione collettiva della conoscenza e dell'identità culturale di un territorio. Stimolare azioni nei confronti della cittadinanza volte a contribuire alla creazione di contenuti culturali permette di generare esternalità positive di grande impatto¹⁸. La conoscenza partecipata può essere una possibile chiave di crescita per i territori attraverso la promozione di un'identità di comunità, ma è necessario che questi processi passino da una mediazione professionale, non certo invasiva, ma storicizzata e sufficientemente filologica, al fine di evitare ricostruzioni finte o del tutto fuorvianti.

¹⁷ F. Valacchi, *Se l'archivio è artificiale. Verso uno ius archivi partecipativo?*, «AIDA Informazioni», 1, 2, 2023, pp. 153-170.

¹⁸ Cfr. P. Feliciati, *La valorizzazione dell'eredità culturale in Italia. Atti del convegno. Supplementi. Appendice 4.*, «Il Capitale Culturale. Studies on the Value of Cultural Heritage», 5, 2016, pp. 205-213; Un esempio efficace di queste esternalità positive è rappresentato dal progetto di valorizzazione dell'Archivio storico del Banco di Napoli: M. Sessa, *L'archivio storico del Banco di Napoli: un patrimonio per la cultura mondiale. Un progetto di valorizzazione e fruizione*, in *L'archivio e le banche: ricerca, tutela, gestione, Atti delle giornate di studio Napoli 11-12 maggio 2000*, a cura di M. Sessa, Luciano, Napoli 2001, pp. 107-104. Sul fronte della restituzione aumentata si veda anche: C. Damiani, *La memoria rappresentata: dalla descrizione inventariale agli archivi narranti*, cit.

In questo senso, il lavoro che il Consorzio della Val D'Orcia tramite il progetto SLIT (Sistema Lettura Integrata del Territorio) sta conducendo può diventare un esempio virtuoso di generazione di saperi, in quanto agisce a partire da quelle testimonianze documentarie e materiali oggetto di studio e rappresentazione nel tempo e nello spazio. A ciò si aggiunge un utilizzo parsimonioso e attento degli strumenti tecnologici, che diventano veicoli di restituzione delle informazioni attraverso la progettazione di sistemi integrati che permettono di spingere i beni "oltre i limiti fisici" che essi naturalmente hanno.

Il fatto non nuovo, ma certamente apprezzabile, è la volontà del Consorzio di partire dal bene, cioè dall'archivio e, con il supporto dei professionisti, sviluppare una narrazione che guidi il pubblico alla scoperta del territorio. Ciò si traduce in un'azione "digitale" che mette al centro l'autorità interpretativa ed evita delle rappresentazioni fuorvianti, soprattutto se rivolte ad un pubblico che diventa parte attiva nel processo di disseminazione della narrazione.

Mentre ci muoviamo inesorabilmente verso un futuro sempre più interconnesso e digitale, dobbiamo continuare a fare attenzione a preservare l'integrità non solo fisica, ma anche scientifica, del patrimonio culturale. Progetti come SLIT pongono sì l'accento su una partecipazione attiva della comunità, ma si concentrano in egual misura anche sulla corretta mediazione professionale e su un utilizzo responsabile delle tecnologie.

Ontologie, web semantico, Intelligenza Artificiale sono strumenti potenti, ma per arrivare ad una valorizzazione del patrimonio significativa – ed edificante sia per chi lavora al progetto sia per chi ne usufruisce – è necessario che ci sia un equilibrio tra innovazione tecnologica e comprensione profonda della materia. Il rischio altrimenti è di cadere nella trappola della superficialità o della mera estetizzazione del contenuto. La vera essenza del patrimonio culturale risiede nel suo contesto, nella sua storia, nelle sue relazioni e nelle persone che lo hanno creato e custodito nel corso degli anni. La tecnologia non deve sostituire o oscurare questa essenza, ma piuttosto la deve amplificare e renderla accessibile a un pubblico più ampio. In tal modo si potrà garantire che le "generazioni future"¹⁹ abbiano la possibilità non solo di vedere e interagire con queste testimonianze del passato, ma anche di comprenderne il vero significato e valore.

In fondo quello che rimane al pubblico del patrimonio culturale sono le storie, le narrazioni. Date, dettagli su eventi, o tecniche di costruzione sono importanti ma non hanno la capacità di raggiungere gli strati profondi della memoria di chi ascolta. Anche per questa ragione il fulcro del lavoro degli archivisti, dei bibliotecari, degli storici dell'arte, degli archeologi risiede nella mediazione tra patrimonio culturale e pubblico, utilizzando tutti i mezzi e gli strumenti a loro disposizione, inclusi quelli tecnologici, senza lasciarsi sopraffare dai *bit* e dai *byte*.

¹⁹ Richiamate spesso nel Piano Nazionale di Ripresa e Resilienza.