

Archivi e Data Visualization. Esplorazione ed esposizione delle fonti mediante la multidimensionalità dei dati digitali e le intelligenze artificiali

«DigItalia» 2-2023
DOI: 10.36181/digitalia-00085

Maria Grazia Berlangieri

Sapienza Università di Roma

L'articolo propone un'interpretazione della digitalizzazione non come esclusiva copia: funzione vicariale del documento monumento (analogico), bensì l'idea di digitale come "riapertura" e "processo". Si illustreranno quindi alcuni esempi di riutilizzo delle fonti mediante l'impiego della Data Visualization e dell'Intelligenza Artificiale. Con la Data Visualization è possibile raccontare un fenomeno fornendo un'immediata visione d'insieme e preservando al tempo stesso la singolarità del dato. Con l'Intelligenza Artificiale si possono mettere a sistema le informazioni esistenti, ma anche generare nuovi dati.

In una contemporaneità che richiede nuovi strumenti interpretativi è essenziale aggiornare le nostre mappe, ovvero riformulare i paradigmi attraverso i quali analizziamo le fonti e la relazione che intrattengono con il complesso fenomeno della digitalizzazione. Rispetto a quello che definirei pluralmente "digitalità" vorrei ritornare al punto di partenza, cioè su cosa significhi digitalizzare e su un tema a mio avviso dirimente: il superamento della concezione di digitale come mera copia del reale.

La digitalizzazione non è un'equazione perfetta il cui risultato è la copia (per quanto molto accurata) della fonte. Questo è un retaggio culturale che abbiamo importato dall'esperienza analogica. Copia infatti rimanda alla funzione vicariale dell'opera/documento "monumento". Tuttavia essa è una delle funzioni possibili – preliminare e certamente necessaria – che ha diversi vantaggi, ma anche dei limiti (intesi nell'accezione positiva di "confini"). Il linguaggio matematico che risiede dietro ogni dato digitale è ontologicamente aperto e manipolabile. Per semplificare, la digitalizzazione avviene tramite il prelievo di campioni (*samples*) da un segnale analogico e continuo nel tempo il cui risultato è un segnale discreto, che viene in seguito quantizzato, convertito, quindi codificato e reso accessibile a qualsiasi elaboratore digitale¹. Come ha scritto Maurizio Ferraris nel suo libro *Documanità*, il fatto che lo statuto dei segni manipolabili non si limiti a numeri e lettere, ma investe suoni, immagini, ovvero che ogni cosa sia documentabile e che questa documentabilità possieda uno standard unificato cambia il mondo con la forza di una guerra in tempo di pace². Vorrei quindi soffermarmi di più sulla "natura" della conversione digitale: per riandare all'etimologia della parola, conversione significa mutare forma o natura delle cose, destinare a un uso diverso. Secondo un neologismo coniato insieme a Flaviano Pizzardi, artista 3D e direttore tecnico con il quale ho collaborato in diversi progetti di ricerca, questa conversione digitale si può ricon-

¹ Teorema del campionamento di Nyquist-Shannon, cfr. <<https://teoriadeisegnali.it/libro/html/libro-4.1.html>>.

² Maurizio Ferraris, *Documanità. Filosofia del mondo nuovo*, Bari-Roma: Editore Laterza, 2021, p. 7.

figurare come *de-sampling*. Cioè come decostruzione del campionamento a favore di un processo combinato la cui "originalità" del campione iniziale può mutare in un'altra forma e addirittura in un nuovo contenuto, quindi cambiare la sua ontologia. Risiede qui l'aspetto più interessante e polisemico delle digitalità: parlo di digitalità al plurale come quell'insieme di pratiche e di oggetti d'arte risultanti da un processo creativo in cui tecnologia, tecnica, narrazione, corpo e politiche culturali agiscono seguendo un paradigma rizomatico che non produce opere uniformi e si caratterizza per una forte convergenza intermediale. Il superamento della copia (del vero, del reale, del documento monumento) avviene attraverso il "montaggio" o "assemblaggio"³. Il tutto e le sue parti (per esempio di un documento, di un'opera d'arte ecc.) si ristrutturano in un processo esposto che ritiene ogni forma possibile, itera e altera la produzione di senso. Le arti digitali hanno dato moltissimi esempi del riassetto dei dati digitali rispetto alle "fonti" originali. Un esempio di questa concezione processuale può venire dall'analisi dell'impiego delle tecnologie digitali nel lavoro artistico di Quayola⁴ o Refik Anadol⁵, i quali a partire dalla digitalizzazione di documenti o opere artistiche "monumentali", "riaprono" le opere digitalizzate per riconfigurarle in un nuovo montaggio, un nuovo assemblaggio, che espone il processo di campionamento e conversione offrendo una metarappresentazione dell'ontologia dei dati digitali, eminentemente cioè quella di essere "aperti e manipolabili".

L'avvento della registrazione digitale e dell'archiviabilità diffusa, addirittura quantistica⁶, spingono l'essere umano a un nuovo passo nella sua naturale relazione con l'artificiale⁷; oggi, ancor più, con l'avanzamento dell'intelligenza artificiale che determinerà nei prossimi anni un'altra cruciale pagina del "digital turn". Nella prima fase della diffusione del digitale, tra la fine degli anni Novanta e il primo decennio degli anni Duemila, lo si riteneva un mero strumento di traduzione del mondo da analogico a digitale secondo il principio di copia e dematerializzazione della fonte analogica. Ogni cosa può essere facilmente campionata e trasformata in un documento digitale, con una potentissima capacità di ricavare plusvalore: cioè qualcosa che eccede per difetto o per eccesso la copia della fonte. I processi di digitalizzazione creano le condizioni tecniche affinché la fonte possa essere riaperta e manipolata⁸. Inoltre le nuove tecnologie e la diffusione della rete web hanno ulteriormente esacerbato quello che

³ In questo caso mi riferisco all'idea di montaggio di Deleuze, vedi Gilles Deleuze, *Cinema 1 – Immagine-movimento*, Milano: Ubulibri, 1984; Id., *Cinema 2 – Immagine-tempo*, Milano: Ubulibri, 1989; ampiamente anticipata in Id., *Differenza e ripetizione*, Bologna: il Mulino, 1971; Id., *Logica del Senso*, Milano: Feltrinelli, 1975; mi riferisco inoltre alla teoria di assemblaggio di Manuel DeLanda, *Assemblage Theory (Speculative Realism)*, Edinburgh: University Press, 2016.

⁴ «Quayola utilizza la tecnologia come lente per esplorare le tensioni e gli equilibri tra forze apparentemente opposte: reale e artificiale, figurativo e astratto, vecchio e nuovo. Costruendo installazioni immersive, spesso in siti architettonici storicamente significativi, Quayola si impegna e reimmagina l'immaginario canonico attraverso la tecnologia contemporanea. La scultura ellenistica, la pittura antica e l'architettura barocca sono alcune delle estetiche storiche che servono come punto di partenza per le composizioni astratte di Quayola. La sua variegata pratica, tutta derivante da software informatici personalizzati, comprende anche performance audiovisive, video, sculture e opere su carta». Cfr. Quayola, <<https://quayola.com/info>>.

⁵ Il media artist Refik Anadol «crea arte pubblica che esplora il potenziale delle intersezioni tra media, scienza e tecnologia per comunicare radicalmente i dati, invitando al contempo il pubblico a visualizzare architetture alternative di spazio e percezione. Attraverso installazioni immersive site-specific, dipinti di dati e performance audiovisive, Refik Anadol Studio prende i dati invisibili che ci circondano come materiale primario e la mente computerizzata come collaboratore, per creare esperienze poetiche». Cfr. Refik Anadol *Studio AI and Architecture*, <<https://refikanadolstudio.com/>>.

⁶ La tecnologia quantistica si basa sui principi fondamentali della fisica e promette di superare i limiti dei tradizionali elaboratori che processano dati tradotti in bit binari (0 o 1).

⁷ Mi riferisco a quella corrente di pensiero che spazia da Arnold Gehlen ad André Leroi-Gourhan, da Walter Benjamin a Pietro Montani, per citarne in maniera parzialissima solo alcuni.

⁸ Maria Grazia Berlangieri, *Performing Space. Lo spazio performativo e l'hacking digitale, nuove tecnologie e transmedialità*, Roma: Bordeaux Edizioni, 2021.

Marshall McLuhan ha definito collasso del rapporto causale⁹ e al punto tale che oggi la “registrazione” precede o convive con l’evento, in quanto la registrazione digitale lo codifica e lo diffonde in realtime sulle piattaforme web¹⁰. Come sottolinea Maurizio Ferraris il web infatti è il più grande apparato di registrazione esistente con miliardi di dispositivi connessi e produce più documenti di qualsiasi fabbrica al mondo¹¹. Byung-Chul Han riprendendo il pensiero di Vilém Flusser¹² parla di «impero neoliberalista della prestazione» in cui i social network rafforzano la costrizione alla comunicazione poiché maggiore comunicazione (digitale) equivale a maggiore capitale¹³. Grazie al fatto che la registrazione digitale utilizza un codice sostanzialmente unico per tradurre immagini, suoni, numeri, lettere, dati biologici, e che le interfacce software sono elaborate affinché gli utenti le utilizzino facilmente e per lo più gratuitamente, ogni persona ha una esponenziale capacità produttiva di documenti (nell’accezione fin qui esposta). I confini tra cosa è materiale e immateriale, naturale e artificiale, sono sempre più sfumati e il tempo che passiamo all’interno delle piattaforme per comunicare, mangiare, acquistare beni di consumo o esperienze di vita, è tale che si è letteralmente fuori dalla storia umana se si è fuori da questa registrazione perpetua della propria esistenza. Questa incessante richiesta di “capitale documentale” ridefinisce il perimetro della rappresentazione del mondo e di noi stessi. Anche in ambito artistico, i surrogati tecnologici, per così dire, competono con le competenze un tempo esclusive del solo “capitale umano”, interferendo con le secolari modalità di rappresentazione. La relazione centripeta e centrifuga tra tecnologie digitali ed esseri umani è caratterizzata in qualche modo da quella che Rancière ha definito una redistribuzione di spazi e di tempi comuni, che sposta il confine “poliziesco” tra chi sa, sa fare, vedere ecc. da quanti non ne sarebbero capaci¹⁴. Ad esserne capaci oggi sono anche le “macchine” stesse e non solo il pubblico; anzi gli spettatori acquisiscono fenomenologicamente competenze proprio per le capacità algoritmiche della macchina e delle sue interfacce complesse: quanto più la macchina rende in grado il pubblico di agire, tanto più registra, e se registra impara creando plusvalore autonomamente. I big data frutto della dataficazione del reale hanno una dimensione tale che possono essere “capitalizzati” da elaboratori con grande capacità di calcolo, mediante gli algoritmi e quindi dalle intelligenze artificiali.

Ma qual è lo strumento che attua la delega tecnica per l’Homo digitus? Lo strumento che ci permette con semplici comandi di interagire con la macchina, (computer, smartphone ecc), è l’interfaccia grafica, la quale organizza e dispone la nostra modalità di accesso al mondo digitalizzato, e attraverso la quale rappresentiamo noi stessi nel mondo, ma anche come “rappresentiamo” per esempio l’apparato documentale, quindi come tramandiamo informazioni e conoscenza. Veniamo a un altro punto, a mio avviso, dirimente: le interfacce grafiche come spazio di riscrittura. In informatica l’interfaccia grafica (anche conosciuta come GUI, dall’inglese *grafical user interfaces*) è un tipo di interfaccia utente che consente l’interazione tra uomo e macchina in modo visuale, cioè utilizzando visualizzazioni grafiche. Però quando parlo di interfaccia estendo il significato ristretto di interfaccia utente al significato generale del termine che indica l’area o la superficie sulla quale due entità qualitativamente differenti si incontrano. L’interfaccia non è un mero strumento tecnologico che ci permette di utilizzare facilmente i nostri dispositivi, ma è lo spazio attraverso il quale oramai editiamo (secondo il linguaggio e le estetiche speci-

⁹ Marshall McLuhan parla di collasso del rapporto causale originato della velocità dei mezzi di comunicazione di massa dell’era elettrica, cfr. Marshall McLuhan, *Gli strumenti del comunicare*, Milano: il Saggiatore, 1964, ripubblicato nel 2015.

¹⁰ Cfr. Bruno Bachimont, *Between Formats and Data: When Communication Becomes Recording*, a cura di A. Romele, E. Terrone, Basingstoke: Palgrave MacMillan, 2018.

¹¹ M. Ferraris, *Documanità. Filosofia del mondo nuovo*, cit., p. 6.

¹² Si veda, tra gli altri, Vilém Flusser, *La cultura dei media* [1997], a cura di A. Borsari, trad. it. di T. Cavallo, Milano: Bruno Mondadori, 2004.

¹³ Byung-Chul Han, *Nello sciame. Visioni del digitale* [2013], Milano: Nottetempo, 2015, p. 48-51.

¹⁴ Cfr. Jacques Rancière, *Lo spettatore emancipato*, Roma: DeriveApprodi, 2018.

fiche delle piattaforme informatiche che utilizziamo) le nostre identità e la rappresentazione di noi negli eventi e nel mondo. In ambito scientifico, le interfacce sono e saranno sempre più le superfici sulle quali riorganizzeremo le fonti estraendo informazioni, producendone di nuove. Gli archivi attraverso il digitale e le interfacce che mediano i metadati, oltre a essere presidi di testimonianza e conservazione, diventano luoghi “generativi” di arte e nuove informazioni, in ogni ambito disciplinare.

Questa rinegoziazione dei termini entro i quali operiamo il lavoro di ricerca nell’ambito delle *digital humanities* ci ha consentito un’evoluzione del lavoro secolare sulle fonti; sfruttando i nuovi spazi di lettura e di scrittura digitale sono possibili nuove interpretazioni delle fonti stesse. Ciò avviene grazie a un uso non strumentale delle interfacce grafiche nel momento in cui andiamo ad analizzare e a valorizzare i patrimoni documentali digitalizzati. Come ha scritto Mitchell Whitelaw, decenni di digitalizzazione hanno reso disponibile una grande quantità di documenti. Eppure la ricerca, l’interfaccia dominante di queste raccolte, non sempre è in grado di rappresentare questa abbondanza. Whitelaw ha definito questa ricerca “ingenerosa”, cioè trattiene le informazioni e richiede una *query*. Invece le *generous interface* sono interfacce generose e navigabili che possono rivelare la scala e la complessità delle collezioni dei patrimoni digitali¹⁵. Sulla scia di quello che Melvin Wevers e Thomas Smits hanno definito “the visual digital turn”¹⁶, nell’ultimo decennio abbiamo visto l’evoluzione di questo approccio promosso da ricerche multidisciplinari in cui la rappresentazione del patrimonio documentale veniva valorizzato dall’intersezione della Data Visualization e dell’Information Design nei processi di esposizione dei dati. Il macro ambiente in cui sono maturate le *generous interface* è quello dell’architettura dell’informazione¹⁷: al singolo dato viene attribuita una forma la cui dimensione e colore lo andranno a connotare con una particolare funzione. Tale funzione sarà attribuita dagli architetti dell’informazione, che possiamo identificare con tutti quegli esperti, dai programmatori ai data scientist, che gestiscono e organizzano dati. Affinché i dati possano assumere una configurazione portatrice di senso è necessario che ci sia un progetto gestito da un team multidisciplinare in cui la figura del designer ha un ruolo chiave. La Data Visualization infatti consente di raccontare un fenomeno fornendo un’immediata visione d’insieme e preservando, allo stesso tempo, la singolarità del dato. La data visualization è il processo di traduzione dei dati in grafici e altri elementi visivi, la presentazione di informazioni quantitative in visualizzazione, la pratica di tradurre le informazioni in un contesto visivo. Di fatto, è anche una parte enorme del processo di analisi dei dati¹⁸. Il termine infografica invece deriva dalla fusione delle parole information e grafica. Un’infografica prevede la selezione di un set di dati utili a illustrare uno specifico fenomeno che veicola avvalendosi dello storytelling. La data visualization è invece più assimilabile a uno strumento mediante il quale un utente può esplorare liberamente i dati, creando delle associazioni. Nella data visualization non viene imposta una storia ma l’utente ha la libertà di selezionare i dati. Tuttavia è la modalità di fruizione che conferma una effettiva differenza fra infografica e data visualization: l’infografica è progettata affinché al suo interno l’utente venga condotto in maniera lineare nella lettura dell’informazione; nella data visualization invece il processo esplorativo lo posiziona in una struttura narrativa aperta¹⁹.

¹⁵ Mitchell Whitelaw, *Generous Interfaces for Digital Cultural Collections*, in «Digital Humanities Quarterly», 9 (2015), n. 1, <<https://mtchl.net/generous-interfaces-for-digital-cultural-collections/>>.

¹⁶ Melvin Wevers — Thomas Smits, *The visual digital turn: Using neural networks to study historical images*, «Digital Scholarship in the Humanities», 35 (2020), n. 1, p. 194-207.

¹⁷ Mara Pometti — Francesco Tissoni, *Comunicare con i dati. L’informazione tra data journalism e data visualization*, Milano: Ledizioni Editore, 2018, p. 20.

¹⁸ Andrea Benedetti, *Data visualization: cos’è, perché funziona, come farla in modo etico*, <<https://www.agendadigitale.eu/cittadinanza-digitale/data-management/data-visualization-cose-perche-funziona-come-farla-in-modo-etico/>>.

¹⁹ Cfr. M. Pometti, F. Tissoni, *Comunicare con i dati. L’informazione tra data journalism e data visualization*, cit., p. 64-67.

Mi limiterò a citare solo alcuni esempi di archivi che utilizzano infografica e data visualization per rendere la complessità del patrimonio trattato. Il Density Design Lab²⁰ (Politecnico di Milano), tra i molti progetti di visualizzazione complessa, ha realizzato la visualizzazione di *Atlante Calvino. Letteratura e Visualizzazione*²¹, una piattaforma web che consente un’immersione visuale nel vasto corpus degli scritti narrativi di Italo Calvino. L’inizio del progetto risale al 2017, con la sua conclusione nel 2020. Durante l’esplorazione è possibile apprezzare una varietà di visualizzazioni organizzate in base alla tipologia di informazione narrata. Inizialmente si presenta una mappa che identifica tre itinerari principali; selezionando le icone si può accedere a nuove sezioni dove le diverse opere vengono raccontate da molteplici punti di vista. Una legenda e un menu laterale agevolano la comprensione dei concetti forniti. In aggiunta, all’interno del sito è presente la sezione denominata Capta la quale fornisce una descrizione delle scelte progettuali, dei metodi e degli strumenti utilizzati. Un altro interessante progetto è *DivineComedy.Digital*²², realizzato per commemorare i settecento anni dalla morte di Dante Alighieri. L’agenzia The Visual Agency²³ ha sviluppato una piattaforma che collega la produzione di opere d’arte ispirate ai temi della Divina Commedia. Utilizzando un algoritmo d’intelligenza artificiale, le opere sono state organizzate in base alle tre Cantiche, ai cerchi, alle cornici e alle sfere, e infine ai singoli canti (ogni canto contiene da tre a sei scene). Seguendo questo schema, gli utenti possono esaminare attentamente le singole scene, analizzare dettagliatamente le opere corrispondenti e ascoltare una voce narrante che recita i versi relativi a quella parte della scena. Inoltre, agli utenti viene chiesto di condividere segnalazioni su altre opere legate al tema, consentendo così nuovi aggiornamenti e integrazioni.

Vorrei quindi condividere due casi studio in cui sono direttamente coinvolta come ricercatrice. Nell’ambito disciplinare delle Performing Arts ho potuto collaborare infatti a *INCOMMON. In praise of community. Shared creativity in arts and politics in Italy (1959-1979)*²⁴: un progetto di ricerca finanziato dal Consiglio europeo per la ricerca (ERC Starting Grant 2015, principal investigator, prof.ssa Annalisa Sacchi) presso l’Università IUAV di Venezia, in collaborazione con Sapienza Università di Roma. INCOMMON si propone come il primo studio di analisi sistematica del campo delle arti performative come risultato della pratica della comunanza, sia teorizzata che sperimentata negli anni ‘60 e ‘70. In particolare, il progetto mira a studiare la storia del “laboratorio Italia” come luogo in cui la controcultura artistica espressa dalle arti performative ha avuto origine in un contesto caratterizzato da una profonda relazione tra filosofia, politica e pratiche rivoluzionarie. L’apparato documentale raccolto e analizzato è stato pubblicato su una piattaforma digitale realizzata dal già citato Density Design Lab. L’esplorazione del patrimonio documentale acquisito nel corso della ricerca rispecchia un paradigma di analisi che sovverte la forma gerarchica dell’archivio “chiuso”, sfruttando le *generous interfaces* secondo le seguenti viste: a) “Performance Remains”, dove i documenti sono come sparsi su un tavolo, cioè visualizzati in uno spazio tridimensionale e possono tracciare percorsi che saltano da una performance all’altra, collegando materiali nel tempo e nello spazio secondo similarità di forma e colore; b) “Recomposition” invece gioca sull’impilamento dei materiali come tecnica generativa. I documenti, trattati come materiale grezzo, generano visioni inaspettate delle performance; c) “The Shape of Community” mostra (visualizzazione per nodi) la rete di relazioni tra persone ed eventi che caratterizzano la comunità artistica e relazionale tracciata nel progetto INCOMMON; d) “Times and Places” è una vista che illustra l’estensione geografica e temporale della comunità di artisti intorno al periodo oggetto dello studio. La visualizzazione mostra i

²⁰ Density Design Lab, <<https://densitydesign.org/>>.

²¹ Atlante Calvino, Letteratura e visualizzazione, <<https://atlantecalvino.unige.ch/>>.

²² Divinecomedy.digital, <<https://divinecomedy.digital/#/>>.

²³ The Visual Agency, <<https://thevisualagency.com/>>.

²⁴ INCOMMON. In praise of community. Shared creativity in arts and politics in Italy (1959-1979), <<https://in-common.org/>>.

luoghi delle comunità artistiche delle avanguardie teatrali dal 1959 al 1979; e) “Catalogue” è una vista che mantiene la concezione tradizionale dei materiali come documenti. I filtri (type of document, performance, venue, city, year, case study, content provider) offrono un modo per stimolare nuove interpretazioni dei documenti come collezioni (Fig.1, Fig.2).

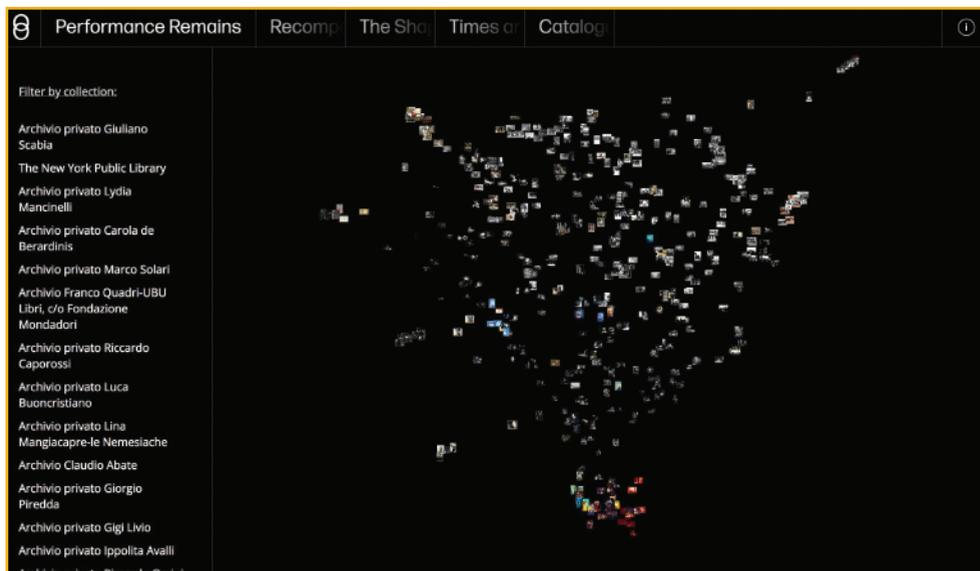


Figura 1. Visualizzazione *Recomposition*, progetto ERC INCOMMON. *In praise of community. Shared creativity in arts and politics in Italy (1959-1979)*, <<https://in-common.org/>>.

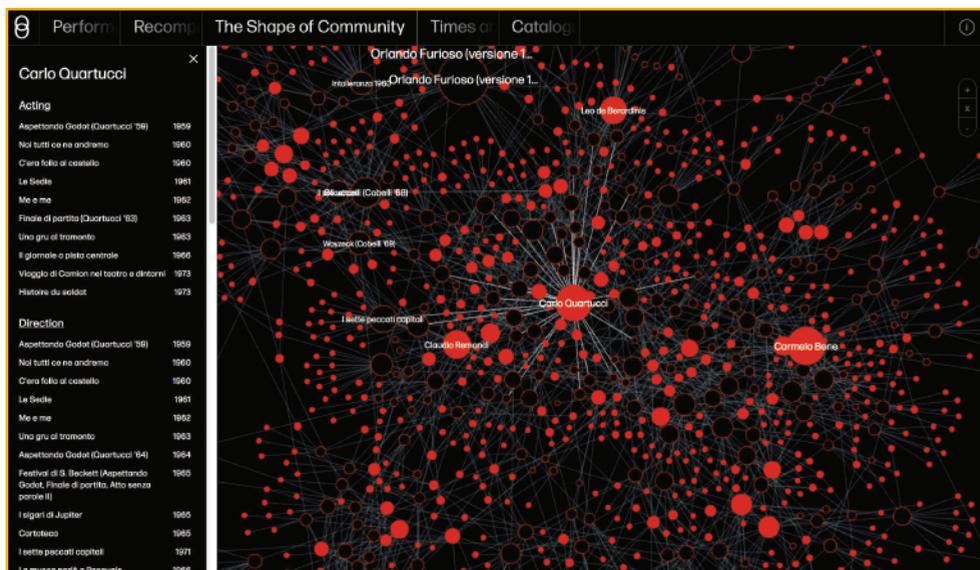


Figura 2. Visualizzazione *“The Shape of Community”*, progetto ERC INCOMMON. *In praise of community. Shared creativity in arts and politics in Italy (1959-1979)*, <<https://in-common.org/>>.

Un caso studio che interseca Data Visualization e Intelligenza artificiale, ancora in fase di completamento, è quello che conduco all'interno del progetto Deep Data Form²⁵, sull'Archivio del Centro Teatro Ateneo (ora Archivio Audiovisivo dello Spettacolo, SARAS, Sapienza Università di Roma). Quest'ultimo nel 2017 è stato riconosciuto dalla Soprintendenza Archivistica e Bibliografica del Lazio di particolare valore storico e culturale; vanta circa 12.000 materiali audiovisivi che documentano prove, seminari e conferenze a cui hanno preso parte le personalità di spicco della cultura teatrale internazionale, dagli anni '30 del Novecento fino ai primi anni Duemila. Attualmente i supporti di archiviazione raccolti nell'archivio sono di natura sia analogica che digitale e troviamo: pellicole 16 e 35 mm, bobine audio per nastro magnetico Nagra, Nastri video da un pollice, D2 (cassette digitali da due pollici), cassette Bvu, U-Matic, Video-8, Betacam SP, Betacam Oxyde, Digital Betacam, Dvcam, mini-DV, Dvd, Dat e infine file DVCPRO, DVCPRO50, Full HD. Anche grazie a un progetto europeo (Eclap, e-library for Performing Arts²⁶) in ambito dei finanziamenti Horizon, sono stati digitalizzati circa tremila contenuti audiovisivi e millecinquecento immagini, fra fotografie e documenti. Per tenere traccia dei contenuti dell'archivio nel corso del tempo era stato redatto un documento inventariale che dagli anni Ottanta fino ai primi del Duemila riportava informazioni in merito all'anno, titolo, regia, collocazione, formato, colore e luogo. Questo foglio inventariale con i suoi metadati è stato oggetto di studio all'interno di una ricerca che ho portato avanti con la preziosa collaborazione della designer Giulia Abbadessa²⁷. Per analizzare e strutturare i dati contenuti nel foglio inventariale dell'Archivio Audiovisivo dello Spettacolo, nelle fasi preliminari del progetto, sono stati utilizzati il foglio elettronico su cui era stato costruito l'inventario dell'archivio, e il software Gephi²⁸ impiegato per la visualizzazione del dataset costruito. Il *dataset building* ha avuto inizio con l'esame dell'Archivio Audiovisivo dello Spettacolo, acquisito sotto forma di foglio elettronico; nella fase iniziale, per avere un'idea generale di tutto l'archivio, sono stati esaminati solo gli items inerenti a Opere e Autori/Autrici, ma affinché potessero essere visualizzati su Gephi era necessario effettuare un adeguamento dei contenuti. È stato creato un documento per identificare i *nodi*, attribuendo un numero identificativo a ciascuna opera e regista e un documento per gli *archi*, in cui i diversi ID venivano messi in relazione²⁹. La prima visualizzazione ottenuta ha messo subito in luce diverse criticità nella catalogazione che presentava refusi, omonimie, incompletezza dei metadati.

Questa tipologia di visualizzazione di un archivio, costruito nell'arco di decenni e con un apparato documentale notevole, ha permesso di far emergere in prima istanza tutta una serie di problemi di inventariazione e catalogazione, quindi si è rivelato uno "strumento diagnostico", considerando l'archivio "come oggetto" nella sua interezza, come opera, ovvero a sua volta documento (Fig. 3, Fig. 4).

²⁵ Dal 2021 coordino come P.I. l'Artificial Intelligence and Machine Learning Laboratory for Digital Humanities, (LABS-SARAS) nell'ambito del progetto "DeepDataForm: Sistema per l'analisi computazionale e il Deep Learning nell'ambito delle Scienze Umane e del Cultural Heritage", finanziato da Sapienza Università di Roma.

²⁶ Eclap, European Collected Library of Artistic Performance, <<https://pro.europeana.eu/project/eclap>>.

²⁷ Giulia Abbadessa, *Archivi e Data Visualization, il caso studio Archivio Audiovisivo dello Spettacolo Sapienza*, tesi di laurea, Corso di Laurea Magistrale interfacoltà in Design, Comunicazione Visiva e Multimediale, a.a.2020-2021, relatrice prof.ssa Maria Grazia Berlangieri, Sapienza, Università degli studi di Roma.

²⁸ Gephi è un software open-source per la visualizzazione e l'analisi di dati. Gephi viene impiegato per analizzare ed esportare dati sotto forma di rappresentazioni grafiche.

²⁹ Per un corretto funzionamento, Gephi necessita di valori inerenti a cosiddetti *nodi* e *archi*. I *nodi* sono gli "atomi" del grafo che rappresentano la natura del singolo dato. Per rappresentare un *nodo*, bisogna attribuire un ID e una LABEL. L'ID è un valore numerico identificativo del nodo, la LABEL invece riporta il nome del contenuto. Gli *archi* collegano i nodi fra loro determinando le relazioni che intercorrono fra i diversi soggetti visualizzati: ad esempio un'opera ed il suo relativo autore (due nodi) saranno connessi da un apposito arco.

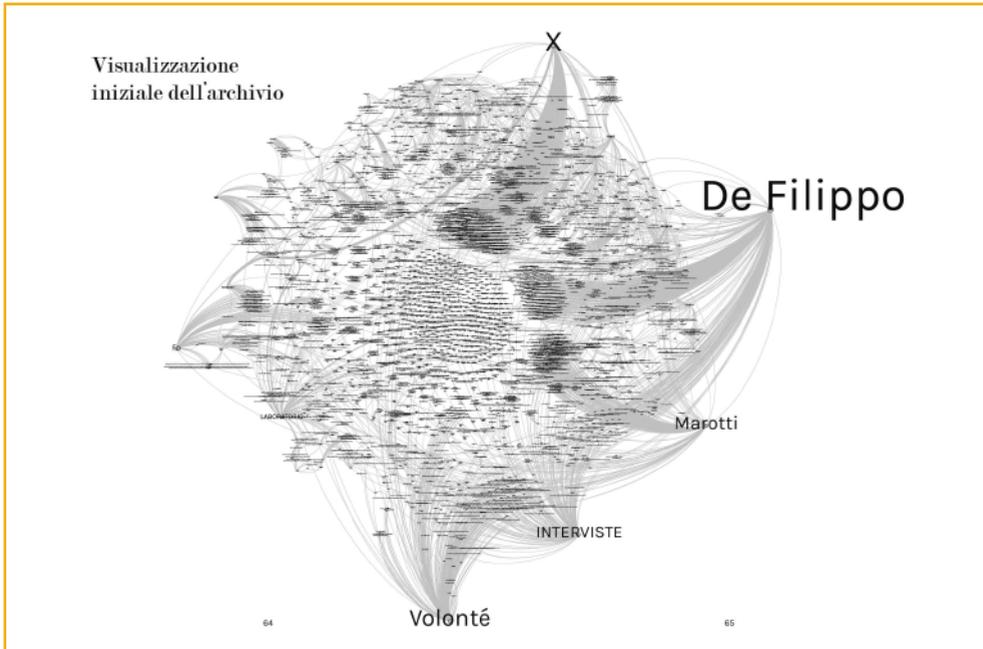


Figura 3. Visualizzazione dell'Archivio del Centro Teatro Ateneo, (ora Archivio Audiovisivo dello Spettacolo, SARAS, Sapienza Università di Roma), a cura di Giulia Abbadessa e Maria Grazia Berlangieri.

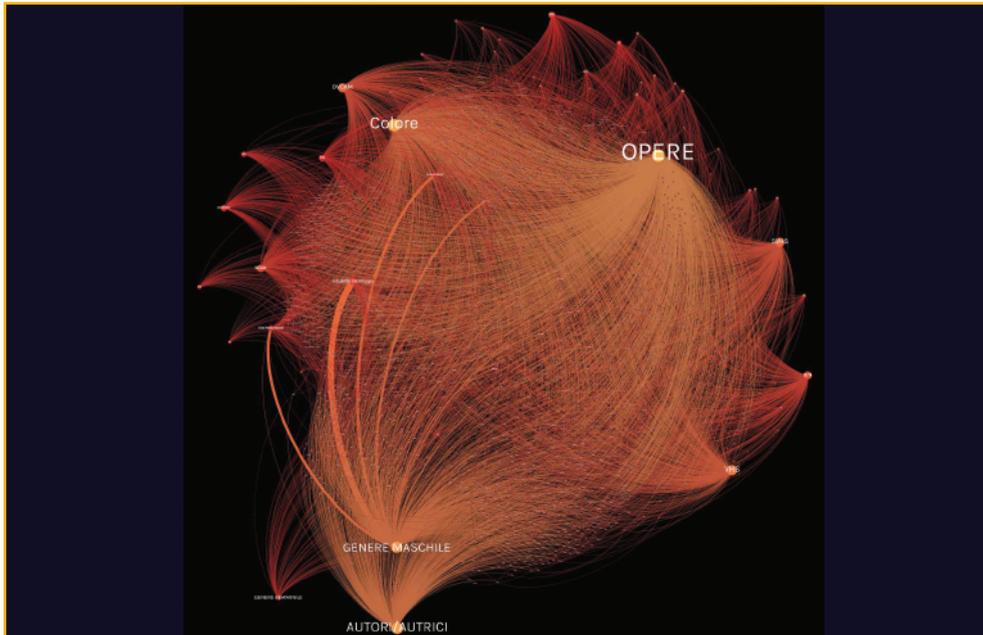


Figura 4. Visualizzazione software Gephi dell'Archivio del Centro Teatro Ateneo, (ora Archivio Audiovisivo dello Spettacolo, SARAS, Sapienza Università di Roma), focus sui documenti audiovisivi di Eduardo De Filippo a cura di Giulia Abbadessa e Maria Grazia Berlangieri.

Secondariamente si è proceduto nell'aggiungere e raffinare le clusterizzazioni, per esempio per tipologia di contenuto (cioè distinguendo tra laboratori, prove, interviste, presentazioni o convegni); inoltre è stata indagata anche la componente di genere dei curatori o curatrici della regia. Questo tipo di operazione è stata effettuata su una porzione di archivio e nello specifico sono state visualizzate 1.300 opere con annessi metadati. Dalle 1.300 opere selezionate all'interno dell'archivio è emerso che hanno collaborato alla realizzazione di quest'ultime 770 artisti, di cui 81 di genere femminile e 718 di genere maschile. Questo dato quantitativo diventa "analitico" attraverso un'unica immagine infografica. (Fig. 5).

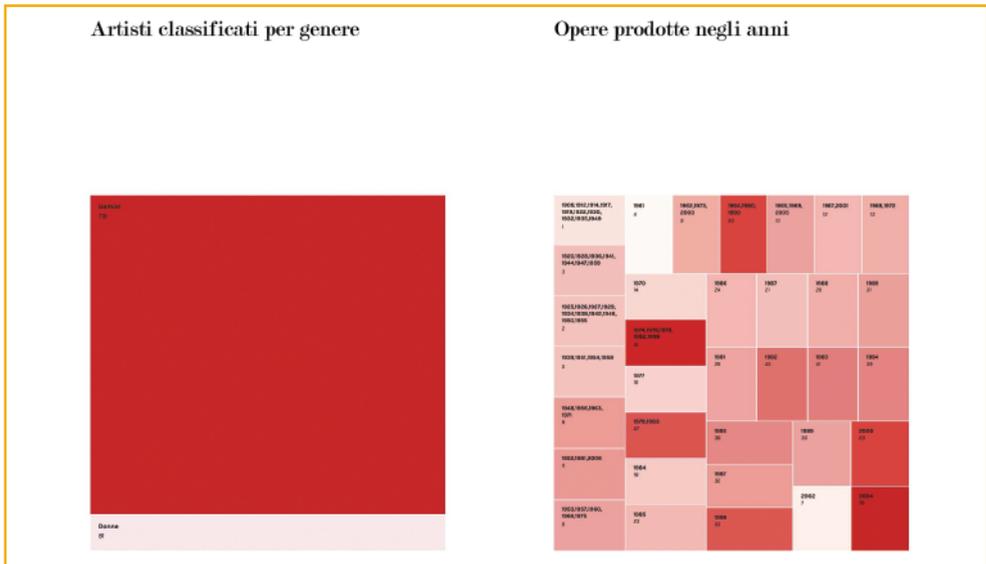


Figura 5. Visualizzazione complessa dell'Archivio del Centro Teatro Ateneo, (ora Archivio Audiovisivo dello Spettacolo, SARAS, Sapienza Università di Roma); focus analitico e quantitativo sulla divisione per generi e clusterizzazione di opere per anni, a cura di Giulia Abbadessa e Maria Grazia Berlangieri.

Lo scopo della mia ricerca attuale è inoltre quello di sperimentare l'utilizzo dell'Intelligenza Artificiale nell'analisi delle sorgenti digitali di immagini e video nel mio ambito disciplinare. Abbiamo quindi deciso di mettere sotto stress la videoregistrazione del Macbeth Horror Suite di Carmelo Bene. Con la designer Giulia Abbadessa si è proceduto quindi a sperimentare l'uso di IA che sfruttano algoritmi di machine learning³⁰. Com'è ormai noto l'Intelligenza Artificiale viene addestrata attraverso enormi quantità di dati a cui uno o più algoritmi permette di esaminare e riconoscere una serie di informazioni, mettendo a confronto una vastissima quantità di input. Per queste ragioni si rivela una risorsa assai efficace e già molto utilizzata, fra le altre cose, nel mondo della visualizzazione degli archivi, come abbiamo visto in alcuni precedenti esempi. Per il caso dello studio sperimentale su Carmelo Bene si è optato per risorse disponibili online quindi un primo software (Azure Video

³⁰ L'intelligenza artificiale è una disciplina appartenente all'informatica che studia i fondamenti teorici, le metodologie e le tecniche che consentono la progettazione di sistemi hardware e sistemi di programmi software capaci di fornire all'elaboratore elettronico prestazioni che, a un osservatore comune, sembrerebbero essere di pertinenza esclusiva dell'intelligenza umana, cfr. Marco Somalvico, *L'intelligenza artificiale*, Milano: Rusconi, 1987. Il Machine Learning (ML) è un sottoinsieme dell'intelligenza artificiale che si occupa di creare sistemi che apprendono o migliorano le performance in base ai dati che utilizzano, cfr. Oracle Italia, *Cos'è il machine learning?*, <<https://www.oracle.com/it/artificial-intelligence/machine-learning/what-is-machine-learning/>>.

Analyzer for Media)³¹ per analizzare la videoregistrazione del Macbeth Horror Suit. In questa sede mi limito a riportare solo alcune delle risultanze ottenute dallo studio. Com'è noto, nel Macbeth Horror Suite ci sono solo due attori, Carmelo Bene nelle vesti di Macbeth e Silvia Pasello in quelle di Lady Macbeth; tuttavia, Azure identifica trentasei personaggi in base alle espressioni facciali rintracciate, nello specifico trenta personaggi riconducibili a Carmelo Bene e sedici riconducibili a Silvia Pasello. I momenti di silenzio vengono ben identificati dal programma, che li registra restituendo una percentuale pari al 42% dell'intera opera. Come ha ben rilevato Giulia Abbadessa, sicuramente un valore così alto rende accreditabile la lettura secondo cui questa palpabile assenza sonora rivesta una funzione narrativa³². Dal punto di vista delle emozioni identificate, la piattaforma Azure è in grado di riconoscerne solamente quattro: gioia, tristezza, rabbia, paura, indicizzandole sulla base delle informazioni audio ricevute, cioè in base a cosa e come viene detto, di conseguenza, in questa particolare opera risulta poco veritiera la risposta ricevuta dal programma. Per intercettare le emozioni ricorrenti nell'opera è stato deciso di analizzarne solo l'aspetto visivo, tralasciando le informazioni sonore. Si è proceduto quindi con estrarre le *thumbnail* dei volti degli attori estratte dal software Azure, le quali a loro volta sono state scremate, eliminando i soggetti mossi o poco identificabili, fino ad arrivare ad un totale di duecento immagini selezionate. Queste, poi, sono state sottoposte al vaglio degli algoritmi di machine learning di Vision AI³³, in grado di rilevare oggetti, volti, emozioni, leggere testo stampato o scritto a mano, individuare la palette cromatica dominante e intercettare il livello di sicurezza di un'immagine (se si tratta di un contenuto violento, per adulti ecc.). Dalle informazioni raccolte emerge una più realistica lettura delle espressioni facciali, tuttavia rimangono perduranti incapacità di lettura delle emozioni della complessa performance di Carmelo Bene. Vorrei però qui sottolineare che perfino dagli errori commessi dall'algoritmo si possono dedurre informazioni sulla per-

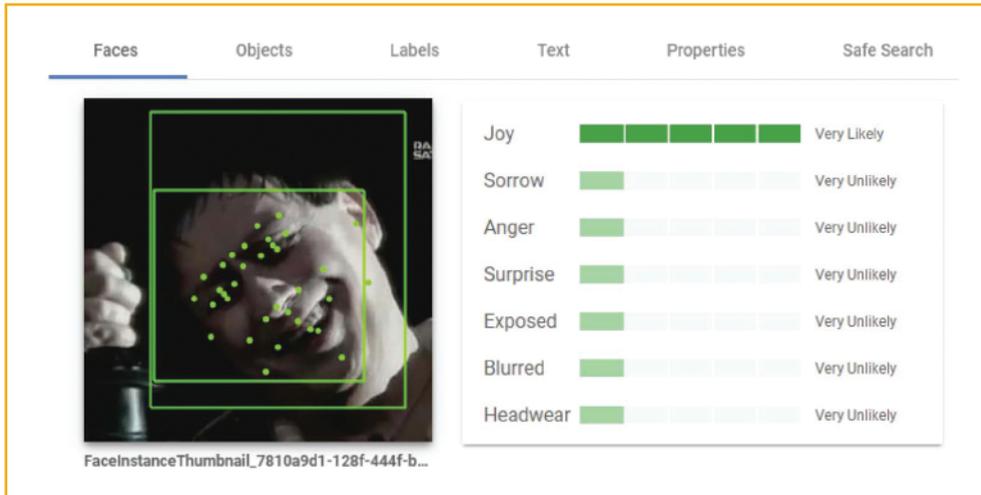


Figura 6. Analisi visiva tramite intelligenza artificiale della videoregistrazione del Macbeth Horror Suite di Carmelo Bene a cura di Giulia Abbadessa e Maria Grazia Berlangieri.

³¹ Azure Video Analyzer for Media è un'applicazione cloud, parte di Azure Applied AI Services, che sfruttando l'intelligenza artificiale (AI) è in grado di ricavare insight da contenuti multimediali.

³² Cfr. Giulia Abbadessa, *Archivi e Data Visualization, il caso studio Archivio Audiovisivo dello Spettacolo Sapienza*, cit.

³³ Vision AI è un ambiente di sviluppo completamente gestito da Vertex AI Vision di Google per creare applicazioni di visione artificiale o ricavare insight da immagini e video con API preaddestrate, AutoML o modelli personalizzati.

formance attoriale. In questa sede mi soffermerò a conclusione di questo articolo su quello che insieme a Giulia Abbadessa abbiamo ritenuto l'errore più "interessante" commesso dall'intelligenza artificiale; cioè Vision AI ha interpretato in un segmento del video le espressioni facciali di Carmelo Bene non come appartenenti a un umano bensì riconducibili a un "animale". In particolare al minuto '51 del video, Macbeth capisce di essere giunto al termine della sua vita, le predizioni delle streghe si stanno avverando; annebbiato da voci maligne provenienti dalla sua stessa armatura il protagonista confessa le sue atrocità e si predispone all'inumano, ovvero alla morte, unica via di uscita da una vita insensata. Naturalmente, più gli algoritmi vengono addestrati, maggiore e più raffinata sarà la capacità di leggere fonti che rispecchiano performance così complesse (tralasciando le opportunità derivanti da una metadatozione automatica) (Fig.6).

In conclusione questo saggio proponendo il superamento del paradigma di digitalizzazione come mera copia, a favore di una visione più complessa che privilegia il concetto di processualità aperta, vuole far emergere le enormi opportunità date dalle nuove tecnologie e al tempo stesso porre l'attenzione sull'importanza delle interfacce grafiche e del "visual turn" nell'editing dei dati digitalizzati. L'ontologia di questi ultimi li rende enormemente soggetti a ricombinazione, quindi soggetti a manipolazione. Questo "editing" dei documenti digitali avviene attraverso le interfacce complesse che unite agli algoritmi di Intelligenza Artificiale possono garantire enormi progressi a livello scientifico, come brevemente ho illustrato nei casi di studio. Tuttavia è un perimetro che gli umanisti in particolare devono presidiare da possibili *bias*, poiché nel *databuilding* si definisce e si definirà sempre di più la produzione di senso. Gli archivi digitali sono chiamati non solo a valorizzare i patrimoni che conservano, ma diventano di fatto "spazi generativi". Nuove sfide attendono "l'archivio", dalla tutela del copyright all'uso dell'Intelligenza Artificiale dei dati e delle immagini immateriali dei nostri patrimoni digitalizzati disponibili online. Il "riuso" dei nostri patrimoni digitali, supera la concezione di copia immateriale delle fonti analogiche. Attraverso il riassetto di algoritmi non è più soltanto una "testimonianza" quella che "fuoriesce" dall'archivio, bensì un riassetto, una rappresentazione. L'eccedenza per difetto o eccesso di questi riassetto determinerà la nuova sfida per gli archivi nell'ennesima pagina del lungo "digital turn". Ancora una volta, Foucault ci viene in soccorso, se si tratta di cosa è enunciabile³⁴. Nell'epoca della disintermediazione e delle macchine intelligenti, l'archivio si trova quindi al centro di macro processi enormi. Aggiornare i paradigmi, non solo negli studi umanistici, sarà oltre che necessario, a mio avviso, "obbligatorio".

The essay suggests interpreting digitisation not as exclusive copy: the vicarious function of the (analogical) monumental document. Rather, it proposes the concept of the digital as a 're-opening' and a 'process'. Therefore, the essay will illustrate some examples of source re-use through the use of Data Visualization and Artificial Intelligence. With Data Visualization it is possible to narrate a phenomenon, providing an immediate overview while preserving the singularity of the data. Artificial Intelligence can organise existing information, but it can also generate new data.

³⁴ Cfr. Michel Foucault, *L'Archéologie du savoir*, Parigi: Éditions Gallimard, 1969.