

# DIGITAL HUMANITIES



Tommaso Mazzoli – Università degli Studi di Trieste

# Digital Humanities: una definizione

- Nel 2010, nel “Manifeste des Digital Humanities”, Mounier definisce le DH come **“una integrazione intensa e a più livelli delle tecnologie digitali in tutti i processi di ricerca, dalla raccolta dei dati fino alla pubblicazione”**
- **1. La svolta digitale intrapresa dalla società modifica e mette in discussione le condizioni di produzione e diffusione della conoscenza.**
- **2. Per noi, le discipline umanistiche digitali riguardano tutte le scienze umane e sociali, le arti e le lettere. Le discipline umanistiche digitali non fanno “tabula rasa” del passato. Al contrario, si basano su tutti i paradigmi, le capacità e la conoscenza specifici di queste discipline, mobilitando gli strumenti e le prospettive particolari del campo digitale.**
- **3. Le Digital Humanities designano una transdisciplina, che porta con sé metodi, dispositivi e prospettive euristiche legate al digitale nel campo delle scienze umane e sociali.**

# Digital Humanities: una definizione

- L'abilità di colei o colui che oggi chiameremmo **digital humanist** è di saper modellizzare l'osservazione di un dominio di interesse in modo tale che la concettualizzazione che ne deriva consenta alla macchina di restituire conoscenza.
- Un atto di interpretazione basato su **procedure, regole e principi** che governano un processo che dai **dati** deve arrivare a restituire **sapere**.
- Il risultato è quello di definire una **nuova metodologia di ricerca**, come esito della riflessione sui metodi e sulle tecniche delle discipline coinvolte: quelle umanistiche e quelle informatiche.
- Se la ricerca umanistica si ripensa alla luce dei sistemi computazionali, analogamente l'informatica affina i suoi metodi acquisendo l'esperienza degli studi sul testo e sul documento.

# Digital Humanities: una definizione

- Si tratta dunque di comprendere cosa succede alle cosiddette scienze umanistiche o “Humanities” quando entrano in contatto con il mondo digitale.
- Da un lato, tutto ciò che è digitale si presenta come un insieme di strumenti messi a disposizione delle scienze umane che se ne servono a tutti i livelli della ricerca, dalla messa in forma alla pubblicazione.
- Dall’altro lato, vuol dire porre alle scienze umane la sfida di comprendere i cambiamenti che la rivoluzione digitale ha comportato e comporterà, sia alle discipline che studiano il mondo che ci circonda sia, più in generale, al mondo stesso.
- Il digitale non fornisce agli umanisti semplicemente dei nuovi mezzi, ma consente loro di porsi domande sui possibili usi degli strumenti digitali a nostra disposizione e sul significato e l’evoluzione di questi stessi strumenti.

# Digital Humanities / Informatica Umanistica

- L'**Informatica Umanistica**, altrimenti chiamata in Italia **Umanistica Digitale**, si confronta già dagli anni 90 con la tradizione della **Humanities Computing** europea e statunitense che poi adotterà la nozione internazionale di Digital Humanities.
- Fino dalle origini la **Humanities Computing** ha rappresentato un'esigenza nata in un contesto di natura **linguistica e letteraria**: manipolare stringhe di caratteri allo scopo di sfruttare la capacità della macchina in termini di rapidità nell'esecuzione delle operazioni e di precisione del risultato.
- Un aspetto meccanico, non ancora metodologico, ma che segnalava l'esigenza di lavorare sui dati testuali, e in particolare sui testi letterari, attraverso sistemi di analisi linguistica puramente quantitativi.
- Linguistica e letteratura, assieme al ruolo della macchina di calcolo come esecutore di algoritmi, sono quindi il punto di partenza necessario per tracciare una cronistoria della disciplina e comprendere come il Web abbia poi modificato in modo significativo, a partire dagli anni 2000, la ricerca di settore.

# Digital Humanities

- L'introduzione dell'informatica o, più precisamente, di macchine per il calcolo automatico negli studi umanistici si può far risalire alla fine degli anni Quaranta, anche se già negli ultimi anni della Seconda guerra mondiale erano stati compiuti tentativi di automazione nel trattamento di dati linguistici per la composizione crittografica di messaggi e la traduzione assistita dai primi rudimentali calcolatori.
- Vedi Alan Turing



- Nato a Vicenza nel 1913, gesuita, Padre Busa è oggi considerato il padre della linguistica computazionale, grazie alla monumentale opera filologica e classificatoria condotta sugli scritti di Tommaso d'Aquino, che ha portato alla creazione dell'*Index Thomisticus*, vale a dire l'indicizzazione completa di tutte le occorrenze di ogni singolo lemma usato nelle opere di Tommaso d'Aquino.
- Il lavoro di Busa oggi rappresenta insieme le origini della Humanities Computing e le nuove tendenze correnti nella gestione di processi di analisi del testo nel **sistema dei big data**, dove si adotta l'analisi statistica di grandi quantità di dati in un'ottica che chiameremo del **distant reading**, ovvero l'utilizzo di metodi computazionali nella gestione dei dati letterari, finalizzati alla scoperta di qualcosa di nuovo.



- Nel lontano 1946, mentre si apprestava a scrivere la sua tesi di laurea alla Pontificia Università Gregoriana (*La terminologia tomistica dell'interiorità. Saggi di metodo per una interpretazione della metafisica della presenza*, Milano 1949), Padre Busa ebbe l'idea di iniziare una indagine accurata sul lessico di Tommaso d'Aquino, a partire dall'intero *corpus* dei suoi scritti.
- Il lavoro si presentava ovviamente monumentale, essendo basato su un repertorio di oltre 9 milioni di parole (contro le appena centomila della Divina Commedia).
- Aveva già compilato **a mano diecimila** schede solo per inventariare la preposizione «in», che egli giudicava fondamentale dal punto di vista filosofico.
- Cercava, senza trovarlo, un modo per mettere in connessione i singoli frammenti del pensiero di San Tommaso per confrontarli con altre fonti.
- In viaggio negli Stati Uniti, padre Busa chiese udienza a Thomas Watson, fondatore dell'Ibm. Il magnate lo ricevette nel suo ufficio di New York. Nell'ascoltare la richiesta del sacerdote italiano, scosse la testa: *Non è possibile far eseguire alle macchine quello che mi sta chiedendo. Lei pretende d'essere più americano di noi.*
- Padre Busa allora estrasse dalla tasca un cartellino trovato su una scrivania, recante il motto della multinazionale: - **Think-** e la frase - **the difficult can be done immediately the impossible takes a little longer**. Lo restituì a Watson con un moto di delusione.
- Il presidente dell'Ibm, punto sul vivo, ribattè: «E va bene, padre. Ci proveremo.»



- Ebbe così inizio una collaborazione duratura che nel 1980 avrebbe portato, dopo quasi due milioni di ore di lavoro, alla creazione dell'Opera Omnia della linguistica tomistica, la madre di tutte le edizioni digitali e ipertestuali a venire.
- Dalle schede perforate ai nastri magnetici, dai CD-ROM ai DVD, fino ad arrivare finalmente sul web nel 2005, il processo di creazione dell'*Index* si è svolto in strettissima connessione con le evoluzioni tecnologiche della seconda metà del XX secolo, costituendo un punto di riferimento costante anche per lo sviluppo della ricerca nel campo dell'Informatica Umanistica europea e mondiale.
- La sinergia tra le intuizioni di un genio visionario made in Italy e il supporto tecnologico di una azienda che ha saputo credere nel valore della ricerca umanistica ha gettato le basi per l'avvento della moderna Informatica Umanistica.
- <http://www.corpusthomisticum.org>

# Digital Humanities

- Il calcolatore elettronico permetteva di considerare e di elaborare quantità di dati che sarebbe stato impensabile dominare con gli strumenti di calcolo tradizionali
- Anche in quest'ambito si sono formati nuovi settori di studio - linguistica quantitativa, statistica linguistica, linguistica matematica, stilometria e lessicometria - con il compito di elaborare e compendiare dati numerici e verbali (statistica descrittiva) ma anche di comprovare l'attendibilità di modelli teorici, soprattutto nello studio di quei fenomeni linguistici che, non potendo essere dimostrati sulla base di un'elaborazione esaustiva dei dati, presuppongono l'impiego di campionature esemplificative.
- Tuttavia, l'accoglienza riservata dagli studiosi alle pubblicazioni che illustravano i risultati di ricerche compiute con metodi matematici e statistici non sempre è stata favorevole.

# Digital Humanities

- Lo sviluppo dei programmi di gestione delle basi di dati è molto legato a due grandi mutamenti: la nascita dell'informatica distribuita e la realizzazione del personal computer.
- Anche se talvolta vengono ritenuti coincidenti per le conseguenze che hanno prodotto, i due fenomeni si sono verificati in tempi successivi, e solo indirettamente possono essere considerati interdipendenti.
- L'affermarsi dell'informatica distribuita dava riscontro all'esigenza sempre crescente di disporre di risorse di calcolo in ambito scientifico, amministrativo e commerciale, e consisteva nella realizzazione di reti capaci di collegare un determinato numero di unità terminali, dislocate nei luoghi di lavoro dei singoli utenti, a un grande elaboratore.
- Solo in un secondo tempo le singole unità terminali sono state dotate di risorse di calcolo autonome via via più consistenti, dando luogo dapprima ai cosiddetti terminali intelligenti e poi a veri e propri elaboratori dedicati all'uso individuale.
- Come sempre è accaduto nella storia dell'informatica, all'innovazione tecnologica dell'hardware ha fatto seguito lo sviluppo o il potenziamento del software di gestione o applicativo, nella spirale di un processo evolutivo che, ingenerando sempre nuove esigenze, ha alimentato un mercato in continuo cambiamento.

# Digital Humanities

- Basi di dati testuali o linguistici (più utilizzate nell'ambito degli studi linguistico-letterari o filologici) e basi di dati fattuali (maggiormente impiegate per il trattamento di dati storici, archeologici o storico-artistici).
- Questa distinzione tende a mettere in evidenza la peculiarità del trattamento di dati più dichiaratamente linguistici rispetto ad altri che presuppongono un intervento descrittivo mediato dall'espressione linguistica: si tratta, in quest'ultimo caso, di acquisire dati relativi a oggetti (monumenti, sculture, dipinti, epigrafi, monete, o altro) o a personaggi, gruppi e classi sociali, avvenimenti storici.
- E' interessante osservare come, nel corso degli ultimi decenni, si sia sviluppato un vivace dibattito sul tipo di linguaggio da utilizzare per la descrizione dei dati fattuali.
- Dalla costituzione di codici per la descrizione delle differenti categorie di oggetti si è passati alla compilazione di vocabolari controllati e, successivamente, alla strutturazione di *thesauri* informatici che consente all'utente di ottenere risposte soddisfacenti anche quando oggetti affini risultino descritti con termini diversi.
- Negli ultimi anni si è registrato un notevole incremento delle ricerche orientate all'identificazione di strutture ontologiche, in vista di una più robusta categorizzazione dei dati, anche multilingui (*Formal ontology*, 1998).

# Digital Humanities

- Intratext <http://www.intratext.com/>
- ALIM Archivio della Latinità Italiana del Medioevo <http://it.alim.unisi.it/>
- Il progetto Codice Pelavicino Digitale <http://pelavicino.labcd.unipi.it/>
- Il progetto di ricerca Musisque Deoque. <http://mizar.unive.it/mqdq/public/>
- Il Libro dei Patriarchi <http://www.librideipatriarchi.it/>
- Umanistica Digitale <https://umanisticadigitale.unibo.it/>