

Digital Humanities e Data Management per i Beni Culturali (2025/2026)

01. Origini

- ▶ Mail: sebastian.barzaghi2@unibo.it
- ▶ ORCID: [0000-0002-0799-1527](https://orcid.org/0000-0002-0799-1527)
- ▶ Sito: sebastian.barzaghi2



Conoscete questa immagine?

Nel 1854 un'epidemia di colera colpisce il quartiere di Soho a Londra, causando decine di morti in pochi giorni.

Il medico John Snow decide di mappare i casi, identificando ogni evento di contagio e morte per colera sotto forma di puntino su una mappa del quartiere.

Questa operazione permette alle autorità cittadine di rintracciare la fonte dell'epidemia fino a una singola pompa dell'acqua.



Cosa intendiamo con *dati*?

- Numeri?
- Statistiche?
- Oggetti digitali?
- Oggetti fisici?
- Software?
- Testi?
- Immagini?
- Video?
- Fonti primarie / secondarie?
- Eventi?
- Interpretazioni?



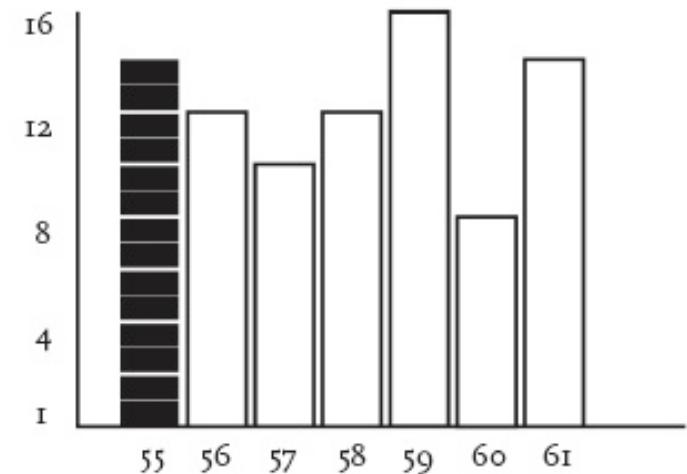
Qualcosa che viene "dato"?

Il numero di nuovi romanzi pubblicati da un singolo editore tra il 1855 e il 1862.

Si tratta di un dato (*datum*, participio passato di *dare*): "qualcosa che viene dato" dalla natura all'osservatore.

Si tratta di un approccio "realista": secondo questa visione, i dati sono indipendenti dall'osservatore.

⚠ Non c'è differenza tra il fenomeno osservato e l'osservazione?

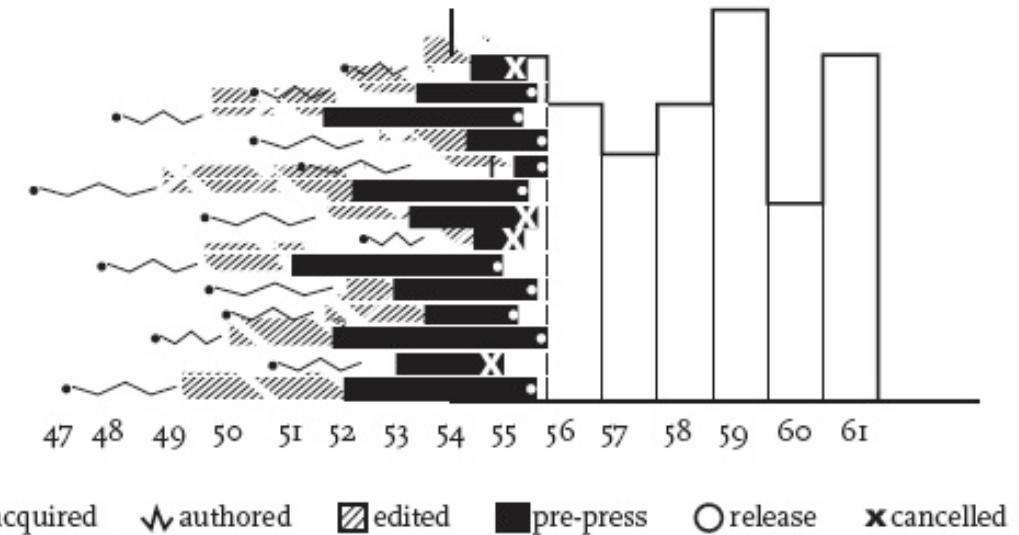


Non è così semplice...

La comparsa nel 1855 di quattordici romanzi dipende da molteplici eventi che si sovrappongono e si influenzano a vicenda: stesura, acquisizione, revisione, pre-stampa e pubblicazione.

⚠ Un dato non esiste di per sé, ma esiste nel momento in cui viene generato, raccolto e/o interpretato attraverso strumenti e sforzi umani.

💡 Ciò che è considerabile *dato* dipende da chi lo usa, da come lo usa, e per quale scopo.

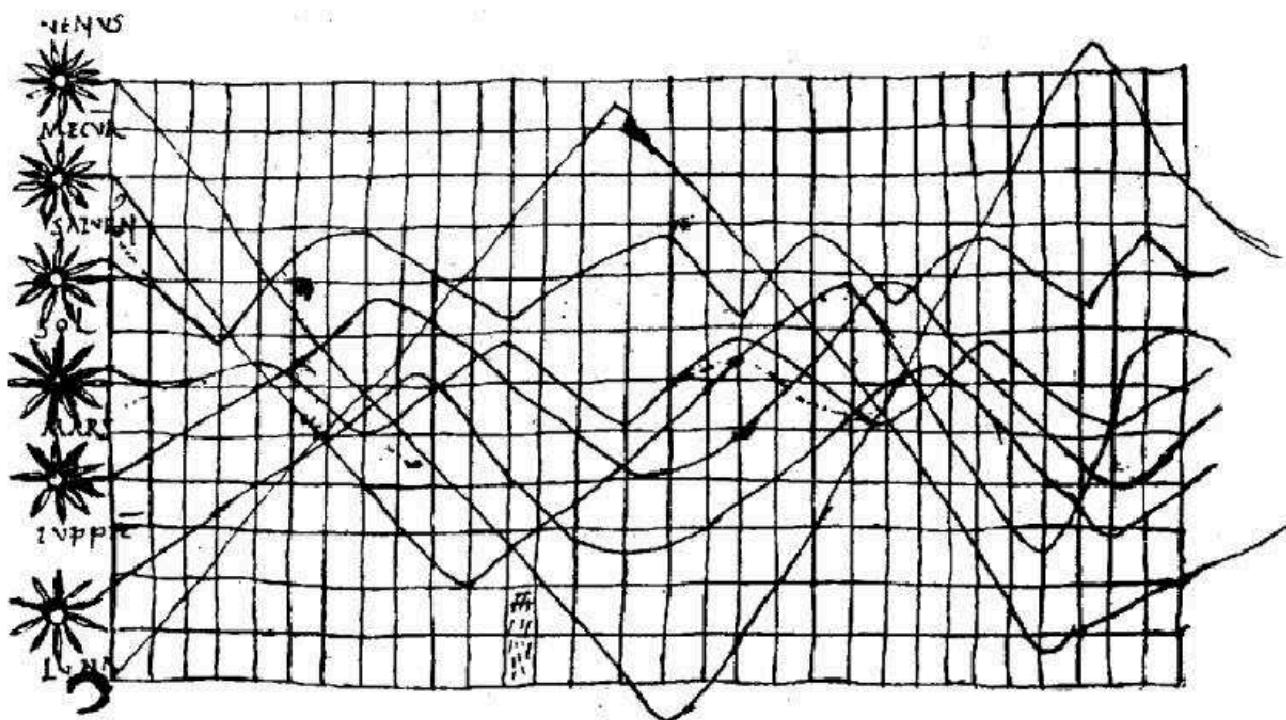


Un dato è "qualcosa che viene preso"

► I dati sono rappresentazioni di entità usate come prova di fenomeni per scopi di ricerca, studio, o altre pratiche di creazione della conoscenza.

Li usiamo come strumenti per comprendere il mondo e modificarlo a seconda delle nostre necessità.

Da sempre sono parte della nostra vita: li utilizziamo, li creiamo, li gestiamo quotidianamente.



Your
visible output

Your
research
data

Publications

Surveys

Bibliographic
references

Translations

Notes

Annotations

Social media
accounts

Interviews

Visualizations

Databases

Images

Quotations

Emails with
colleagues

Corpora

Audio

Video

Timelines

Feedback

Artwork

Experimental data

Project
websites

I dati sono raccolti in dataset

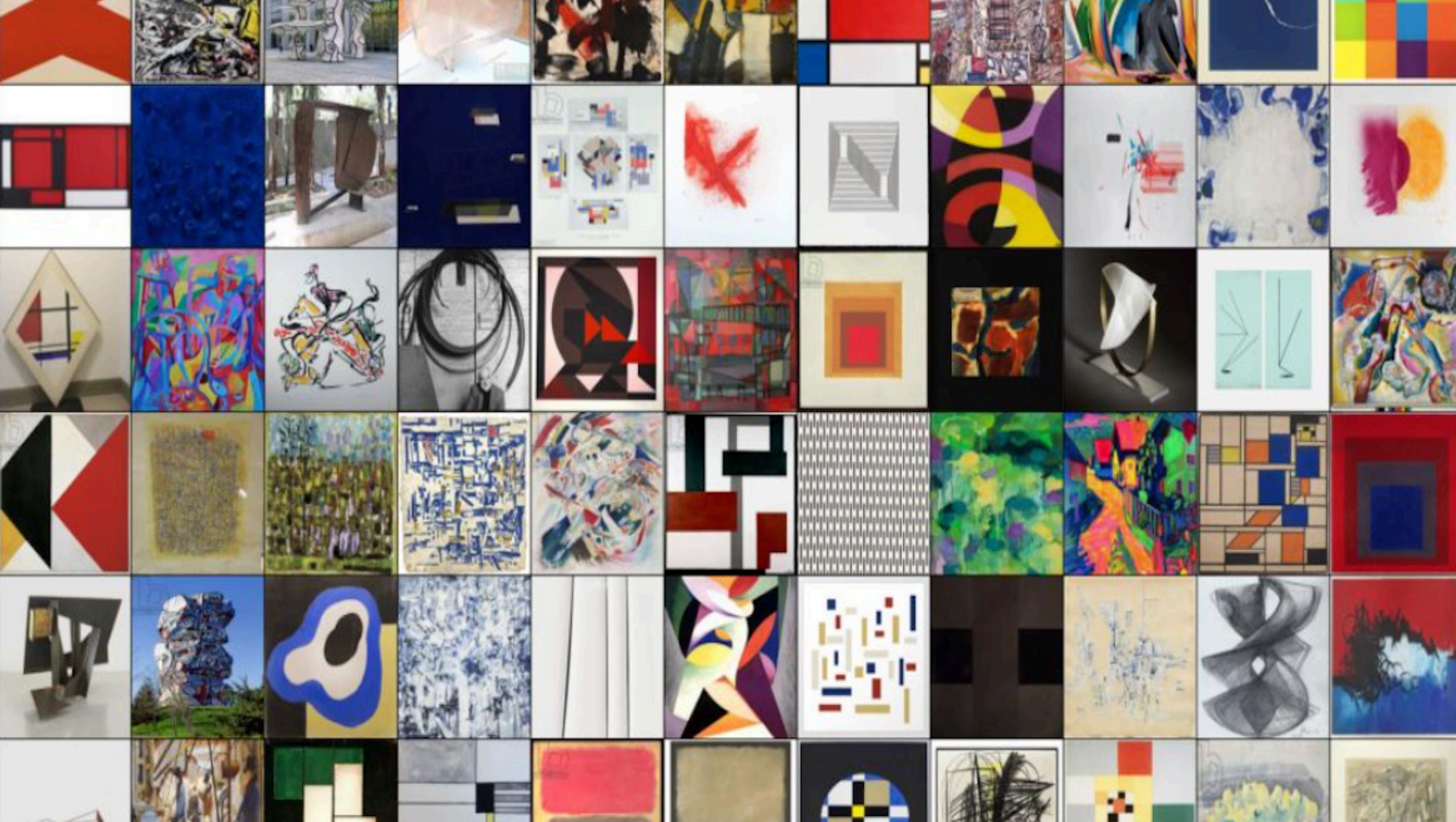
Difficilmente si ragiona in termini di un singolo dato. I dati sono molto più utili se considerati in gruppi, che normalmente chiamiamo *dataset*.

→ Un **dataset** è *una raccolta di dati organizzati secondo criteri precisi*.

⚠ Riflette le circostanze che hanno portato alla sua creazione e gestione (strumenti, limiti, bias, responsabilità...).

🔍 Quanto sono importanti i dataset? [1], [2], [3], [4]





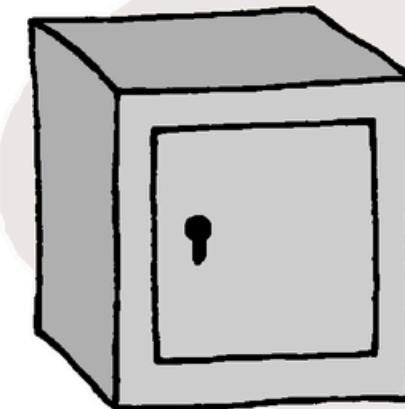
Determinare il valore dei dati

Per valutare l'utilità, il valore e la qualità dei dataset e dei dati che li costituiscono, dobbiamo comprendere il *contesto* nel quale questi dati esistono.

Il contesto è fornito dalla *documentazione* (quando c'è...).

💡 Un ruolo fondamentale, da questo punto di vista, è ricoperto dai **metadati** che documentano *in maniera strutturata* i *chi, i cosa, i come, i quando, i dove, i perché* dei dati.

DATA



METADATA



 Dataedo /cartoon

Piotr@Dataedo

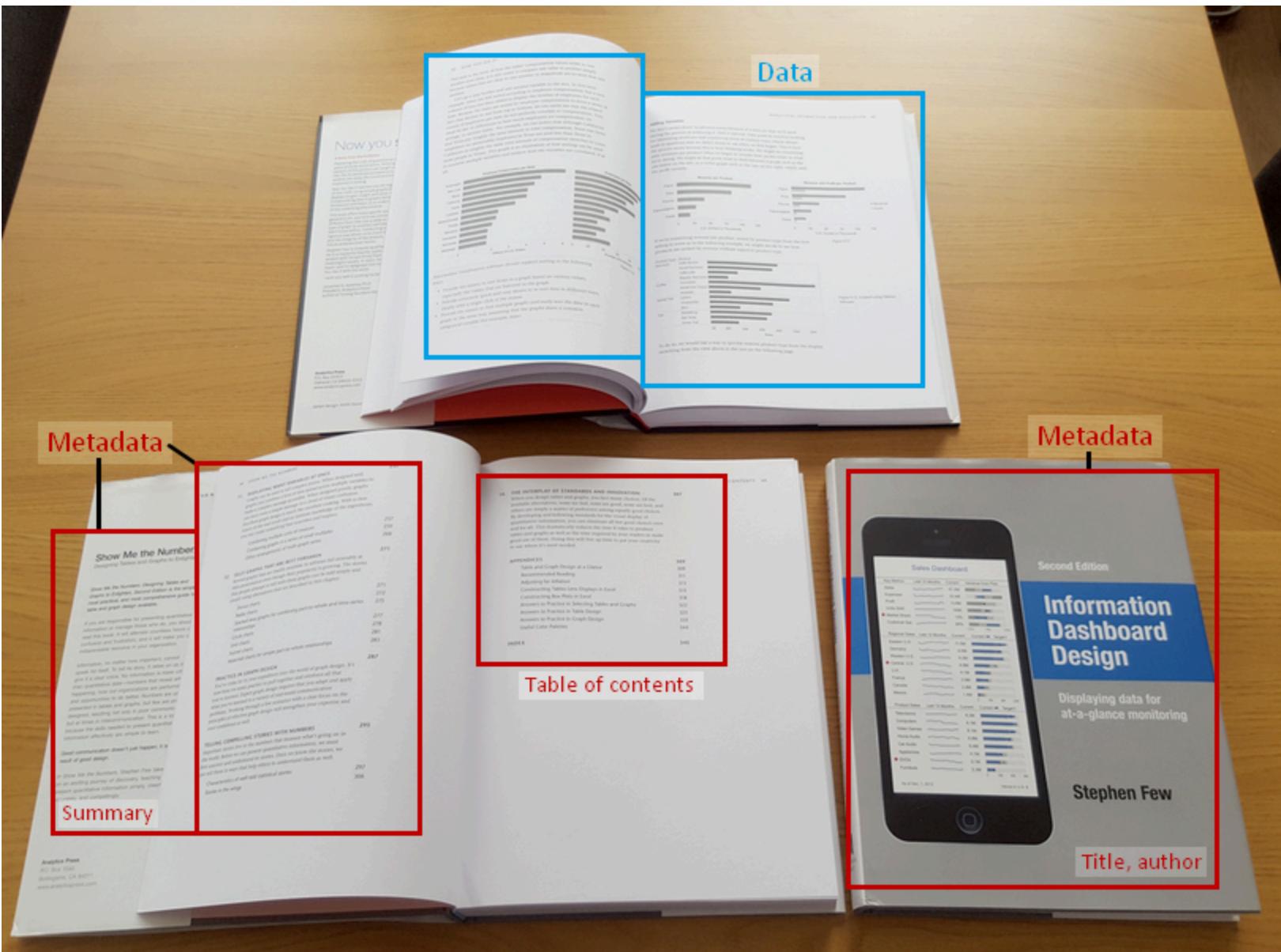


Data

Filename:	Tadzik.jpg
Author:	Piotr Kononow
Date:	August 15, 2016 6:40:10PM
	5,312 × 2,988 JPEG
File:	15.9 megapixels 3,393,448 bytes (3.2 megabytes)
Camera:	Samsung SM-G920F
	4.3 mm
Lens:	Max aperture f/1.9 (shot wide open) Auto exposure Program AE
Exposure:	1/402 sec f/1.9 ISO 40
Flash:	none

A satellite map showing a residential area with streets labeled in Polish. A green arrow points to a specific location on one of the streets.

Metadata



2016 Sales Metadata

Month	Forecast	Sales	Variation
Jan 17	42,000	38,532	-3,468
Feb 17	45,000	41,934	-3,066
Mar 17	45,000	42,163	-2,837
Apr 17	45,000	43,050	-1,950
May 17	45,000	45,145	145
Jun 17	48,000	47,745	-255
Jul 17	48,000	49,623	1,623
Aug 17	48,000	52,539	4,539
Sep 17	45,000	47,324	2,324
Oct 17	45,000	44,700	-300
Nov 17	42,000	44,923	
Dec 17	48,000	51,120	
	546,000	548,798	

Data

James:
Forecast

Dati su dati

→ I **metadati** sono *dati che descrivono un oggetto informativo e ne documentano alcune caratteristiche.*

Esempi di metadati includono l'autore di un libro, il contenuto testuale, la data di pubblicazione, e così via.

💡 I metadati possono essere visti come l'insieme di tutto ciò che si può dire su un dato (oggetto informativo) in un determinato momento, a qualsiasi livello di aggregazione.

Month	Forecast	Sales	Variation
Jan 17	42,000	38,532	-3,468
Feb 17	45,000	41,934	-3,066
Mar 17	45,000	42,163	-2,837
Apr 17	45,000	43,050	-1,950
May 17	45,000	45,145	145
Jun 17	48,000	47,745	-255
Jul 17	48,000	49,623	1,623
Aug 17	48,000	52,539	4,539
Sep 17	45,000	47,324	2,324
Oct 17	45,000	44,700	-300
Nov 17	42,000	44,923	
Dec 17	48,000	51,120	
	546,000	548,798	

I metadati possono *riferirsi* a tante cose

I metadati costituiscono una forma di documentazione (di solito strutturata) che descrive o rappresenta le caratteristiche di un oggetto.

💡 Quindi, possono essere usati per descrivere tanto gli oggetti *fisici* (per esempio, frammenti di vasi, reperti archeologici, quadri, libri, persone, ecc.), quanto gli oggetti *digitali* (per esempio, software, foto, dataset, videogiochi, modelli 3D, ecc.).



I metadati possono essere tante cose

I metadati possono assumere molte forme diverse.

 A partire da semplice testo libero (per esempio, l'etichetta di un'opera d'arte esposta in un museo), pensato per un utilizzatore umano, i metadati possono esistere anche sotto forma di contenuto codificato in un formato altamente strutturato, standardizzato e comprensibile dalle macchine (per esempio, un file MARC).

```
<leader>00000cam a2200000Ma 4500</leader>
<controlfield tag="001">957351941</controlfield>
<controlfield tag="008">051202t19951990enkb g b a001 0 eng d</controlfield>
<datafield tag="040" ind1=" " ind2=" ">
  <subfield code="a">MYIIU</subfield>
  <subfield code="b">eng</subfield>
  <subfield code="c">MYIIU</subfield>
  <subfield code="d">OCLCO</subfield>
  <subfield code="d">OCLCF</subfield>
  <subfield code="d">OCLCQ</subfield>
</datafield>
<datafield tag="043" ind1=" " ind2=" ">
  <subfield code="a">a-ii---</subfield>
</datafield>
<datafield tag="050" ind1="1" ind2="4">
  <subfield code="a">DS436</subfield>
</datafield>
<datafield tag="100" ind1="1" ind2=" ">
  <subfield code="a">Kulke, Hermann.</subfield>
</datafield>
<datafield tag="245" ind1="1" ind2="2">
  <subfield code="a">A history of India /</subfield>
  <subfield code="c">Hermann Kulke and Dietmar Rothermund.</subfield>
</datafield>
<datafield tag="250" ind1=" " ind2=" ">
  <subfield code="a">Rev, updated ed.</subfield>
</datafield>
<datafield tag="260" ind1=" " ind2=" ">
  <subfield code="a">London ;</subfield>
  <subfield code="a">New York :</subfield>
  <subfield code="b">Routledge,</subfield>
  <subfield code="c">repr. 1995, 1990.</subfield>
</datafield>
<datafield tag="300" ind1=" " ind2=" ">
  <subfield code="a">[v], 415 pages :</subfield>
  <subfield code="b">maps ;</subfield>
  <subfield code="c">22 cm</subfield>
</datafield>
<datafield tag="336" ind1=" " ind2=" ">
  <subfield code="a">text</subfield>
  <subfield code="b">txt</subfield>
  <subfield code="2">rdacontent</subfield>
</datafield>
<datafield tag="337" ind1=" " ind2=" ">
  <subfield code="a">unmediated</subfield>
  <subfield code="b">n</subfield>
  <subfield code="2">rdamedia</subfield>
```

I metadati sono sempre associati ai dati che descrivono

I metadati accompagnano ciò che descrivono in vari modi.

- 💡 Possono essere presenti direttamente nei dati stessi (per esempio, un file PDF o un'immagine digitale), oppure possono essere registrati in oggetti separati collegati in qualche maniera ai dati che descrivono (per esempio, file di testo, fogli di calcolo, ma anche una scheda di catalogo o l'etichetta di un'opera esposta in un museo).



Esistono diversi tipi di metadati

- **Descrittivi:** descrivono dati usando dettagli sul contenuto che permettono a questi dati di essere rintracciabili, identificabili e selezionabili (es. titolo);
- **Strutturali:** forniscono informazioni sulla costruzione e sull'organizzazione dei dati all'interno della loro collezione (es. indice);
- **Amministrativi:** permettono la gestione dei dati nel tempo, fornendo informazioni sulle loro origini e le loro modalità di utilizzo (es. licenza).

Item specifics

Condition	Brand New: A new, unread, unused book in perfect condition with no missing or damaged pages. See the ... Read more	ISBN	9780262046305
Publication Year	2022	Type	Textbook
Format	Hardcover	Language	English
Publication Name	Introduction to Algorithms, Fourth Edition	Item Height	2.2in
Author	Ronald L. Rivest, Charles E. Leiserson, Thomas H. Cormen, Clifford Stein	Item Length	9.3in
Publisher	MIT Press	Item Width	8.4in
Item Weight	78.6 Oz	Number of Pages	1312 Pages

About this product

Product Information

A comprehensive update of the leading algorithms text, with new material on matchings in bipartite graphs, online algorithms, machine learning, and other topics. Some books on algorithms are rigorous but incomplete; others cover masses of material but lack rigor. Introduction to Algorithms uniquely combines rigor and comprehensiveness. It covers a broad range of algorithms in depth, yet makes their design and analysis accessible to all levels of readers, with self-contained chapters and algorithms in pseudocode. Since the publication of the first edition, Introduction to Algorithms has become the leading algorithms text in universities worldwide as well as the standard reference for professionals. This fourth edition has been updated throughout. New for the fourth edition New chapters on matchings in bipartite graphs, online algorithms, and machine learningNew material on topics including solving recurrence equations, hash tables, potential functions, and suffix arrays140 new exercises and 22 new problemsReader feedback-informed improvements to old problemsClearer, more personal, and gender-neutral writing styleColor added to improve visual presentationNotes, bibliography, and index updated to reflect developments in the fieldWebsite with new supplementary material Warning- Avoid counterfeit copies of Introduction to Algorithms by buying only from reputable retailers. Counterfeit and pirated copies are incomplete and contain errors.

Product Identifiers

Publisher	MIT Press
ISBN-10	026204630x
ISBN-13	9780262046305
eBay Product ID (ePID)	18050031277

Product Key Features

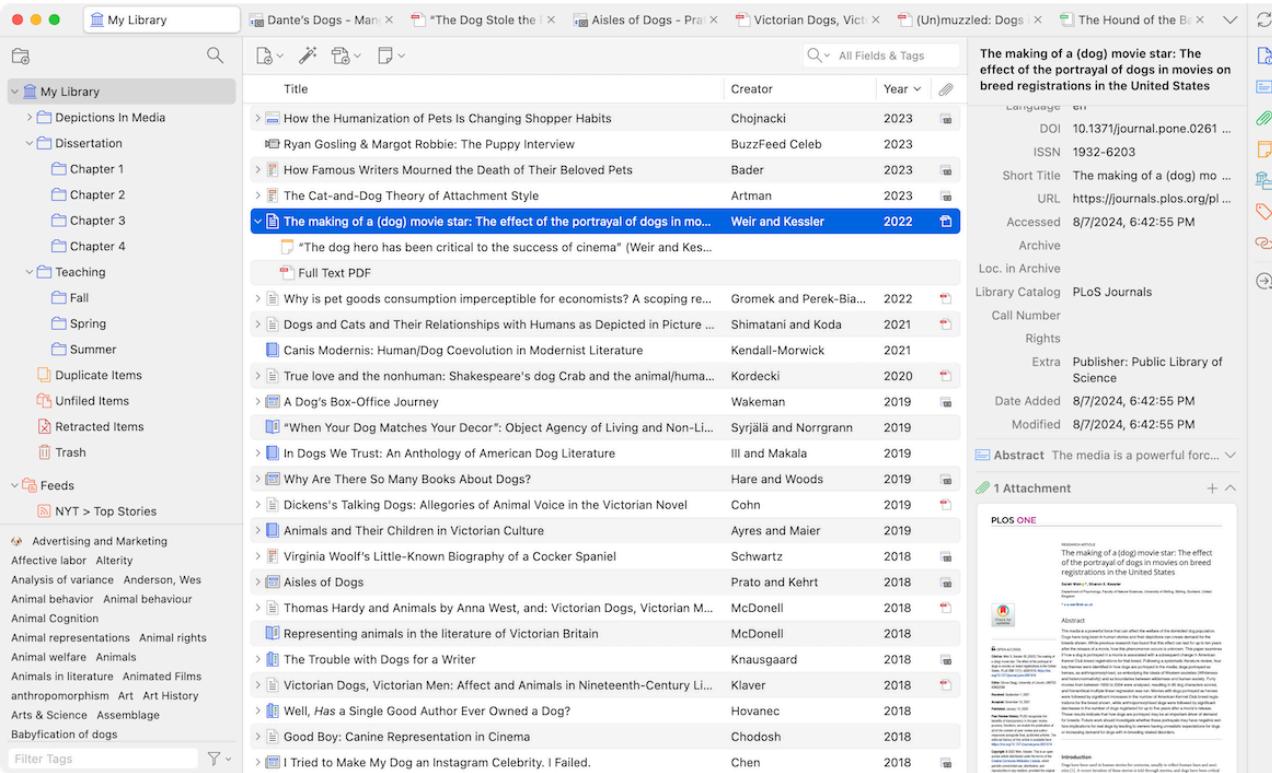
Author	Ronald L. Rivest, Charles E. Leiserson, Thomas H. Cormen, Clifford Stein
--------	--

Anche gli umanisti usano e producono (meta)dati

Ogni tipo di attività basata sulla conoscenza comporta la produzione e/o il riutilizzo di dati, anche se il termine *dato* non viene utilizzato in ogni contesto.

🔍 Esempio: uno studente che utilizza delle fonti (primarie o secondarie) per corroborare i contenuti della propria tesi di laurea produce *almeno* una bibliografia.

🔍 Esempio: [1]



The screenshot shows a library management software interface with a sidebar and a main search results area. The sidebar on the left contains sections for 'My Library' (with sub-items like 'Depictions In Media', 'Dissertation', 'Teaching', 'Feeds', and 'NYT > Top Stories'), 'Advertising and Marketing', 'Affective labor', 'Alterity', 'Analysis of variance', 'Anderson, Wes', 'Animal behavior', 'Animal behaviour', 'Animal Cognition', 'Animal representations', 'Animal rights', 'Animal welfare', 'Animals', 'Animals in literature', 'Animated Films', 'anthropomorphism', 'Art', 'Art History', 'Arts & Science', 'Assemblage', and 'Babyfication of dogs'. A 'Filter Tags' dropdown is also present. The main area displays a list of search results with columns for 'Title', 'Creator', 'Year', and 'Language'. One result is highlighted: 'The making of a (dog) movie star: The effect of the portrayal of dogs in movies on breed registrations in the United States' by Weir and Kessler, published in 2022. The right side of the interface shows detailed information about this article, including its DOI (10.1371/journal.pone.02611), ISSN (1932-6203), short title ('The making of a (dog) mo...'), URL (https://journals.plos.org/pl...), and various metadata fields like 'Archive', 'Loc. in Archive', 'Library Catalog', 'PLOS Journals', 'Call Number', 'Rights', 'Extra', 'Date Added', 'Modified', and 'Abstract'. A preview of the PDF is also shown on the right.

Le discipline umanistiche sono dominate da paradigmi tradizionali

- ⚠ Il modo in cui il linguaggio viene usato nelle discipline umanistiche tende ad essere ermetico, individualista, ambiguo.
- ⚠ I dati umanistici sono spesso non accessibili a causa di licenze troppo stringenti.
- ⚠ I dati umanistici sono complessi e raramente documentati nell'ottica di una loro gestione adeguata.

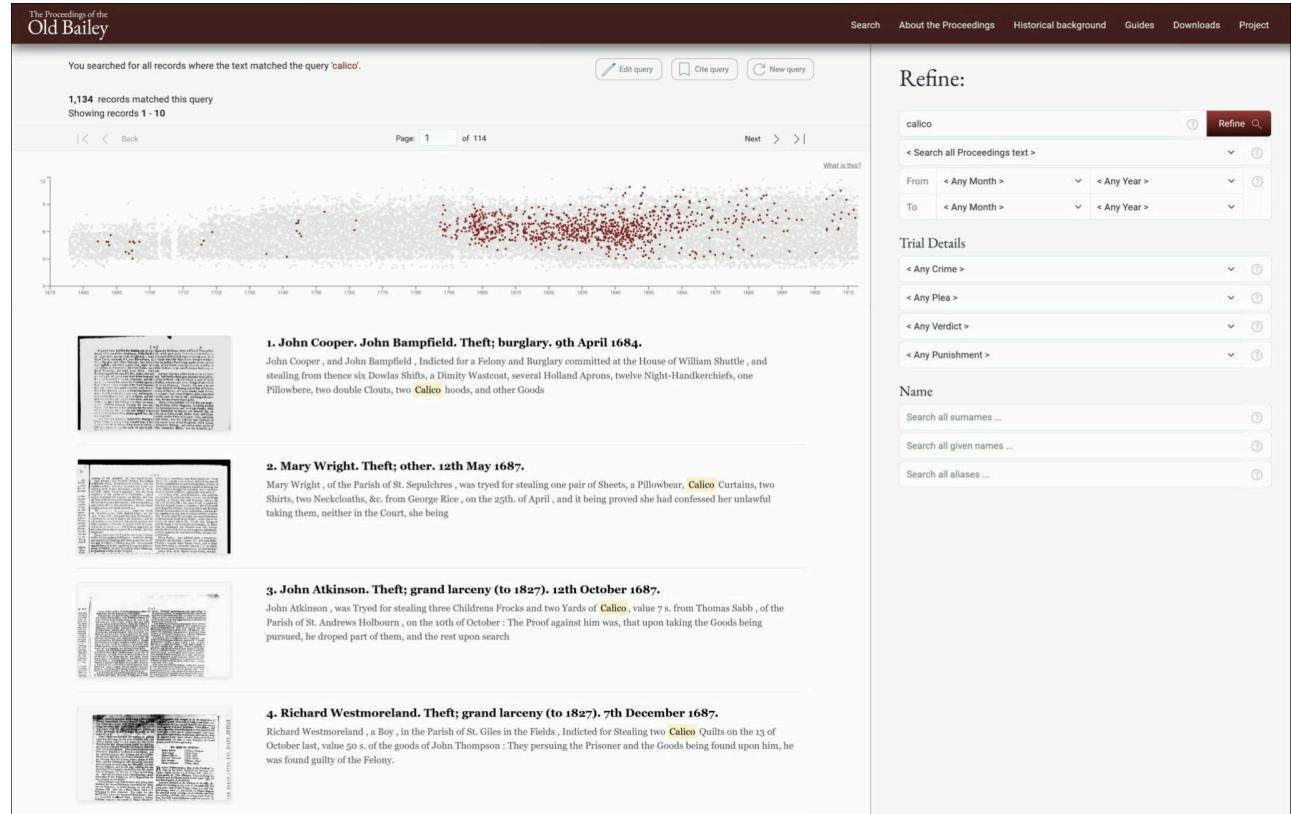


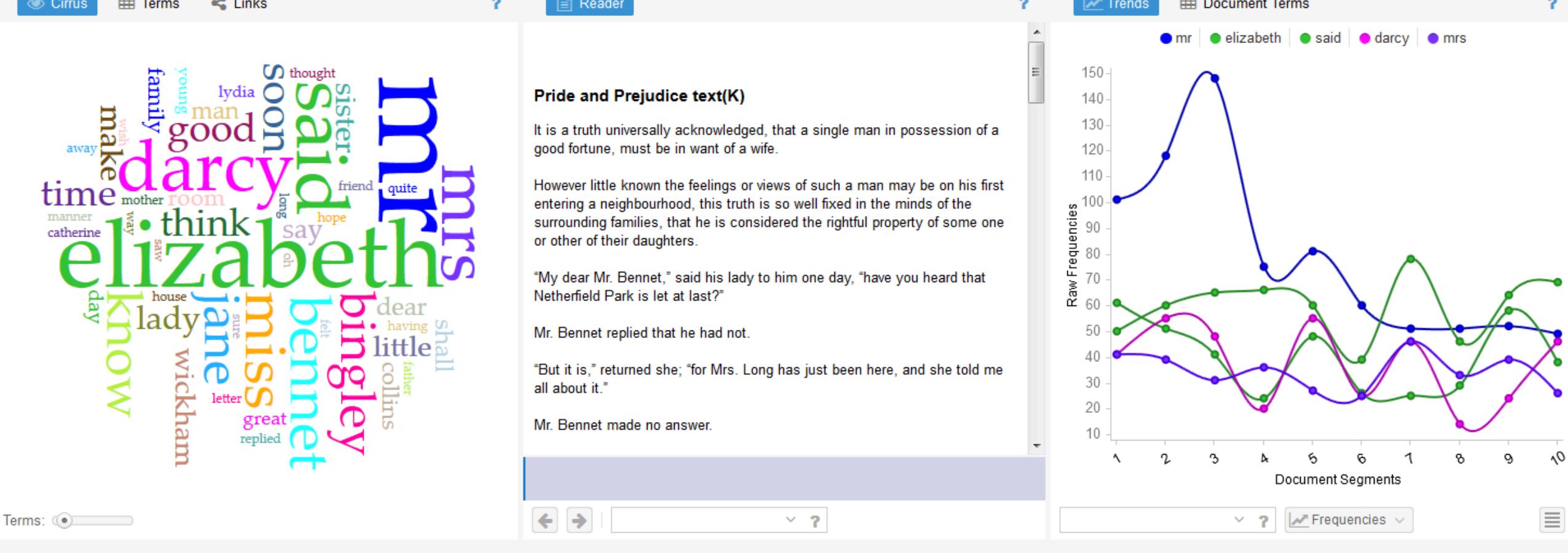
La tradizione si (in/s)contra con il digitale

Le tecnologie digitali mettono in discussione paradigmi esistenti e permettono di sviluppare nuovi metodi di indagine, produzione e disseminazione scientifica dei dati umanistici.

Un paio di esempi:

- [Are Pop Lyrics Getting More Repetitive?](#)
- [Mining the Dispatch](#)
- [The Old Bailey Proceedings](#)





Terms:

Summary Documents Phrases

This corpus has 1 document with 122,077 total words and 6,347 unique word forms. Created 2 seconds ago.

Most frequent words in the corpus: mr (786); elizabeth (597); said (401); darcy (374); mrs (343)

?

Contexts Bubblelines

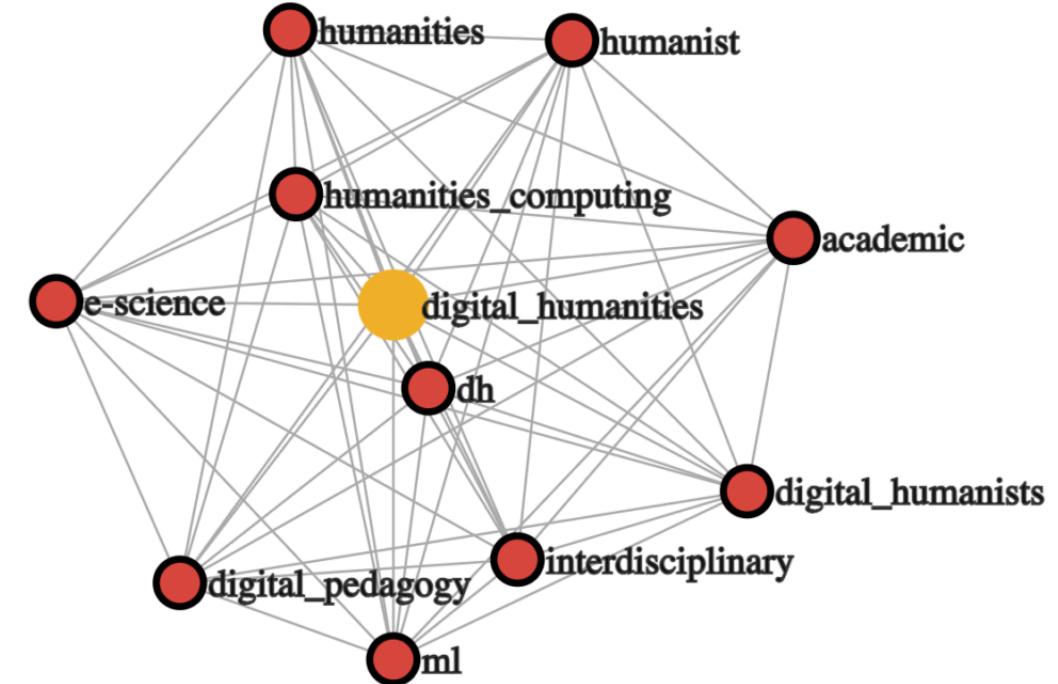
Document	Left	Term	Right
[+] 1) Pride...	of their daughters. "My dear	mr	. Bennet," said his lady to
[+] 1) Pride...	Park is let at last?"	mr	. Bennet replied that he had
[+] 1) Pride...	told me all about it."	mr	. Bennet made no answer. "Do
[+] 1) Pride...	it, that he agreed with	mr	. Morris immediately; that he is
[+] 1) Pride...	it affect them?" "My dear	mr	. Bennet," replied his wife, "how
[+] 1) Pride...	handsome as any of them,	mr	. Bingley may like you the
[+] 1) Pride...	must indeed go and see	mr	. Bingley when he comes into
[+] 1) Pride...	scrupulous, surely. I dare say	mr	. Bingley will be very glad
[+] 1) Pride...	of quickness than her sisters."	mr	. Bennet, how can you abuse
[+] 1) Pride...	I will visit them all."	mr	Bennet was so odd a

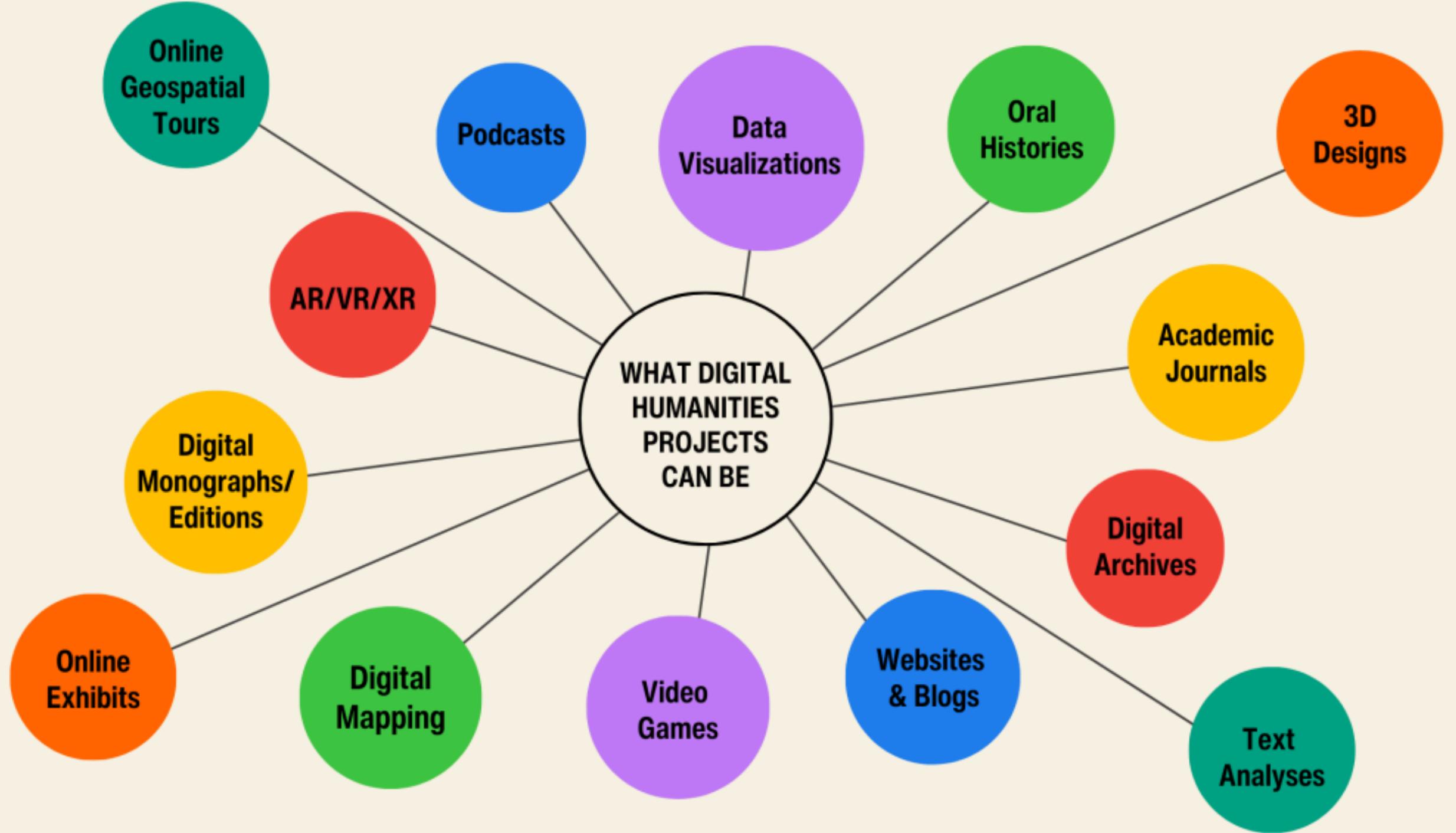
Un nuovo paradigma

→ Col termine **Digital Humanities** (DH) ci si riferisce ad *un insieme di discipline in rapida espansione che esplora, sviluppa e applica metodi e strumenti digitali alle scienze umanistiche.*

Spesso vengono utilizzate tecnologie digitali per esplorare domande di ricerca in ambito umanistico.

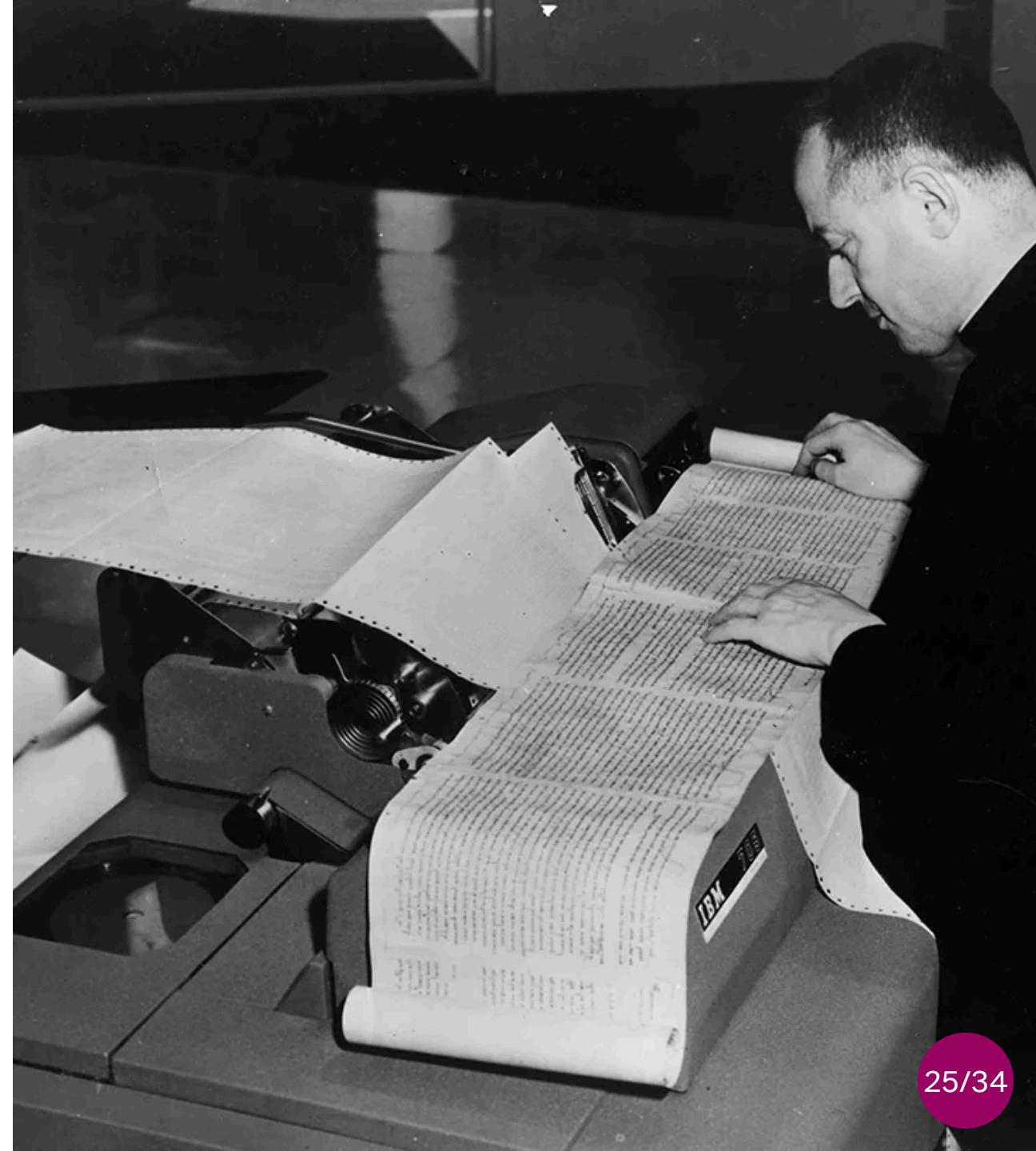
In altri casi, vengono formulare domande di ricerca di natura umanistica su questioni riguardanti tecnologie digitali e il loro impatto.





Le diverse prospettive delle DH

- 1** Le DH nascono da una prospettiva *testuale*, focalizzata sulla digitalizzazione di fonti fisiche e sull'utilizzo di strumenti computazionali per organizzare ed analizzare materiali digitalizzati.
- 2** In seguito, le tipologie di fonti si moltiplicano (immagini, audio, video, ecc.), entrano in gioco anche i materiali nativamente digitali, e cresce un maggiore interesse verso l'epistemologia (il *come*).
- 3** Ora, sta aumentando l'interesse verso l'impatto che la *digitalizzazione*, la *dataficatione*, l'*algoritmizzazione* e l'*AI* stanno avendo sulla società e sulla cultura.

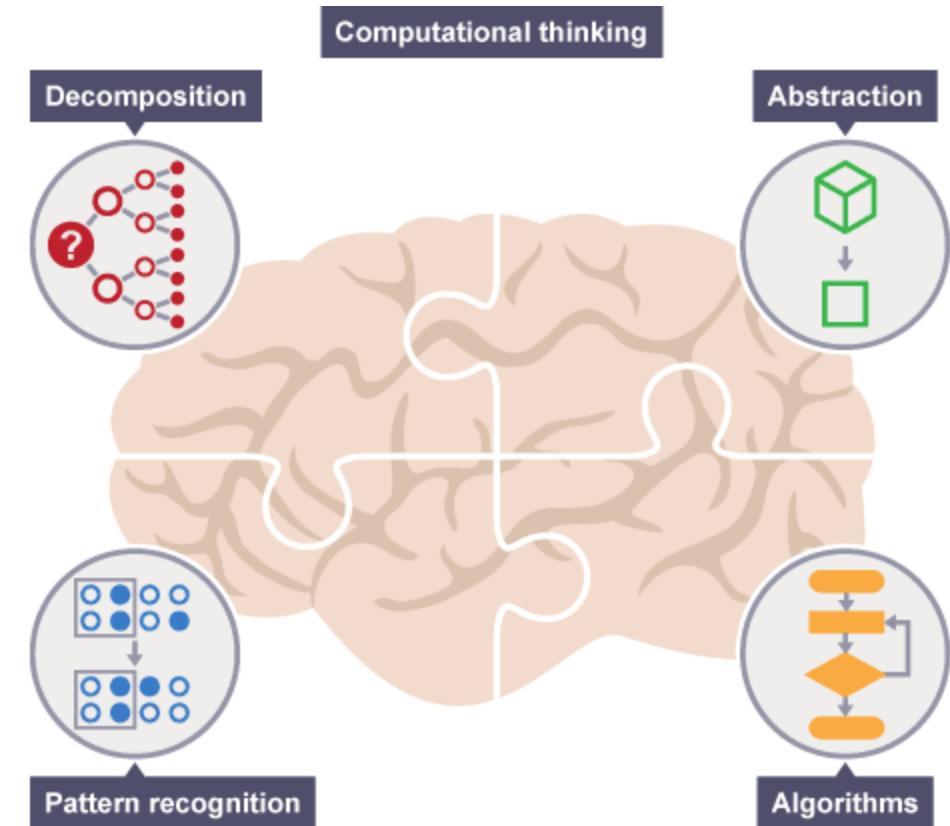


I concetti computazionali chiave

- **Decomposizione:** dividere un problema in parti più piccole e gestibili;
- **Riconoscimento di pattern:** cercare somiglianze e tendenze nei componenti del problema;
- **Astrazione:** focalizzarsi solo sulle informazioni rilevanti, rimuovendo dettagli non necessari;
- **Algoritmi:** sviluppare una soluzione sequenziale al problema.



[Slide su Computational Thinking](#)



Il tempo dei macroscopi

→ Per **macroscopio** ci si riferisce ad *un dispositivo (fisico, digitale, concettuale) che fornisce una visione d'insieme su qualcosa, sintetizzando elementi in relazione l'uno con l'altro e individuando pattern, tendenze, casi particolari, eccezioni, dettagli.*

Ci permette di osservare ciò che è troppo grande, lento, o complesso per essere compreso dal solo occhio umano.

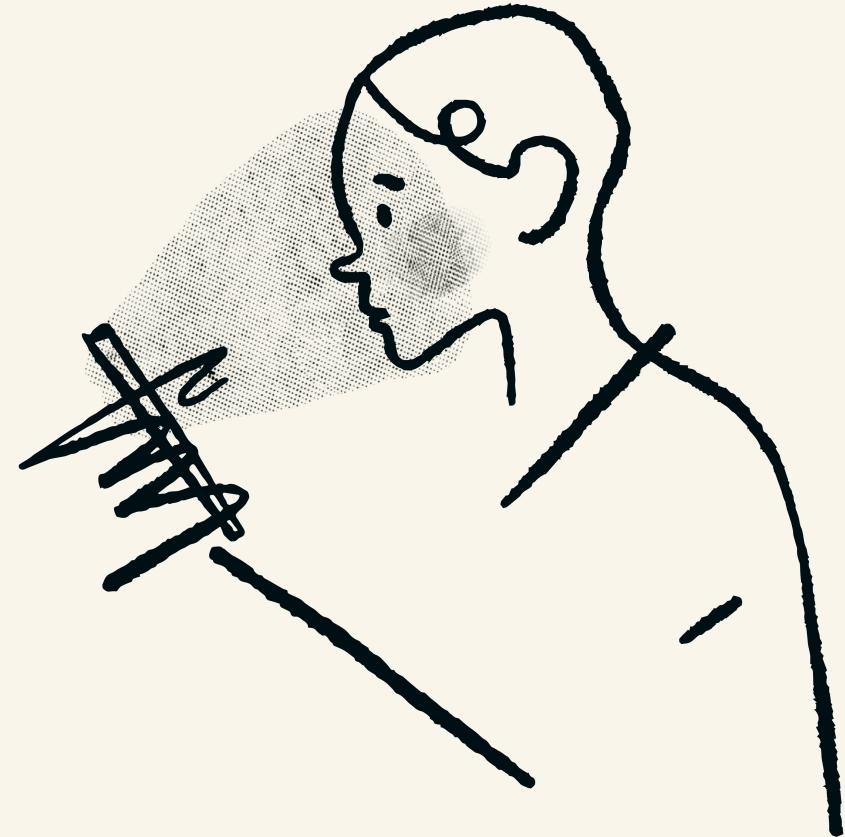
Può essere un dataset, un servizio, un software, una metodologia...



Le DH sono un'emana^{zione} della Digital Literacy

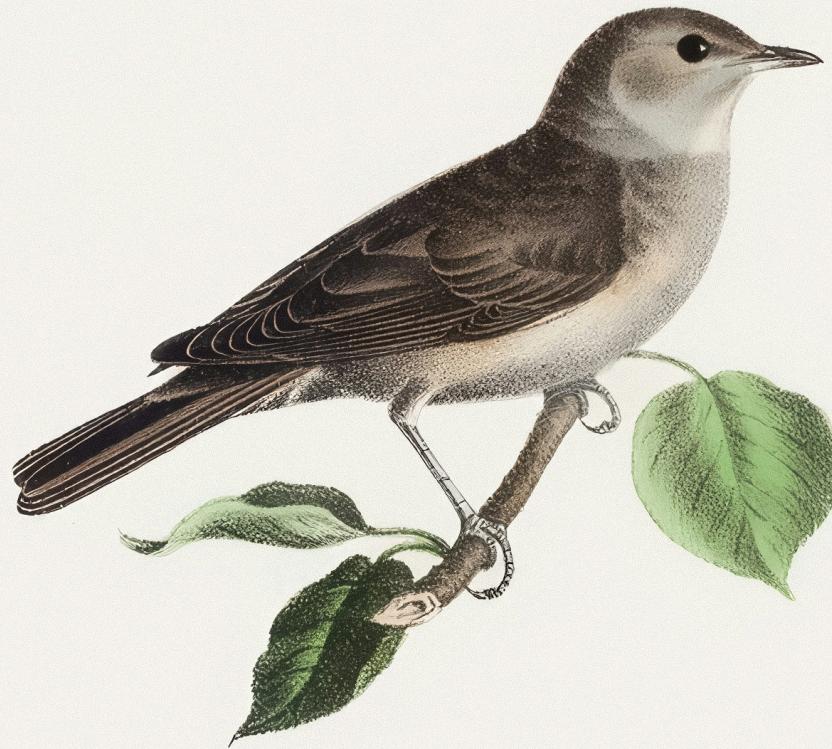
Opportunità di esplorare, sperimentare, e sviluppare abilità trasversali e desiderabili come:

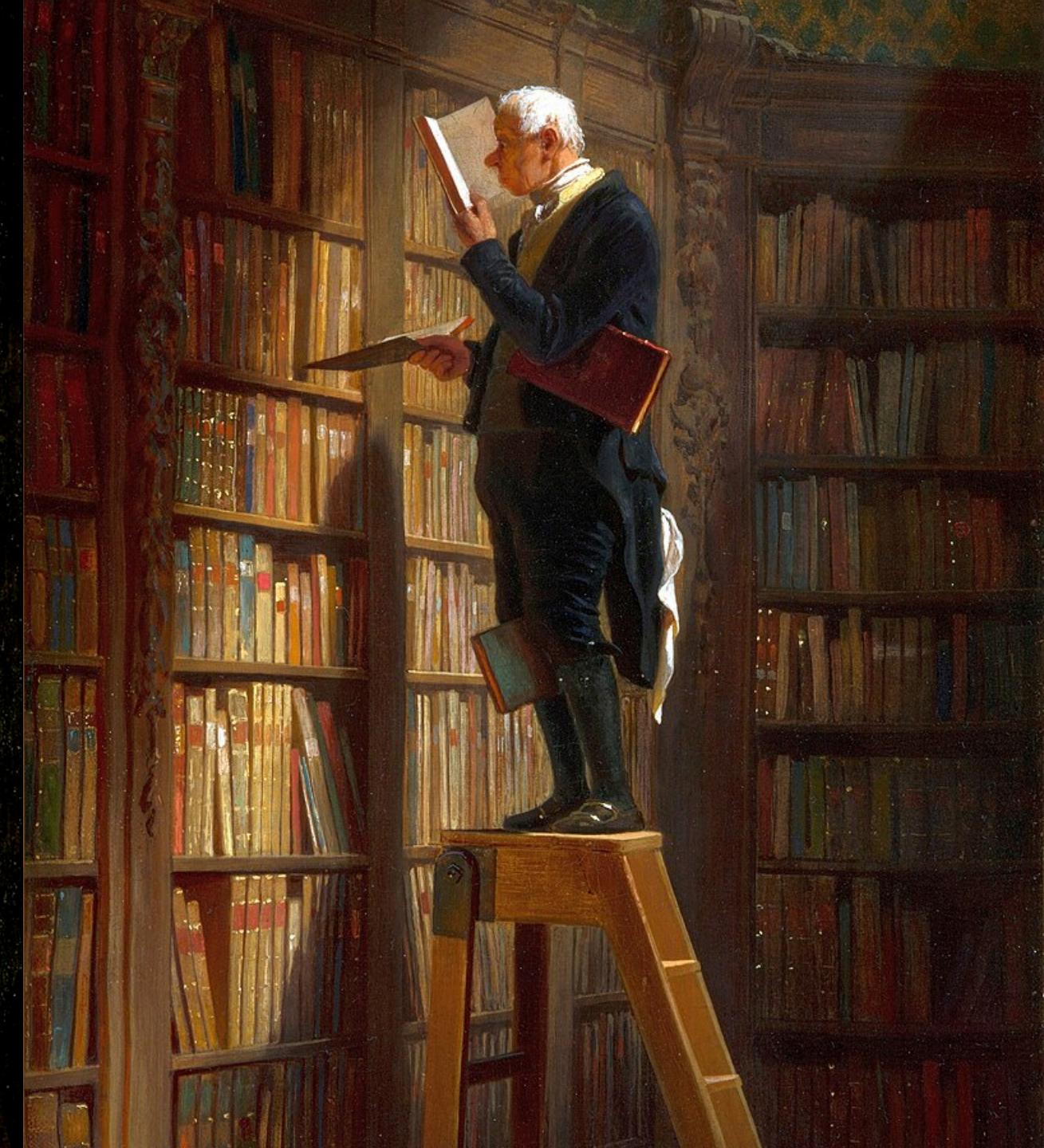
- Pensiero computazionale;
- Gestione dei dati;
- Programmazione;
- Cooperazione interdisciplinare;
- Conoscenza su privacy dei dati, licenze, interazione con l'ecosistema Web e con le AI...



Il progetto d'esame

Spiegazione preliminare





Dataset storico-artistico: MyPOD

Descrizione

Dataset costituito da metadati descrittivi di dipinti che trattano di *mitologia classica*.

Opportunità

Implementazione di un progetto di **gestione dei dati** che coinvolge:

- 1 o 2 domande di ricerca;
- creazione di un database;
- un' **analisi dei dati culturali** sulla base delle domande scelte;
- presentazione del processo e risultati.



Dataset bibliografico: LiSPOD

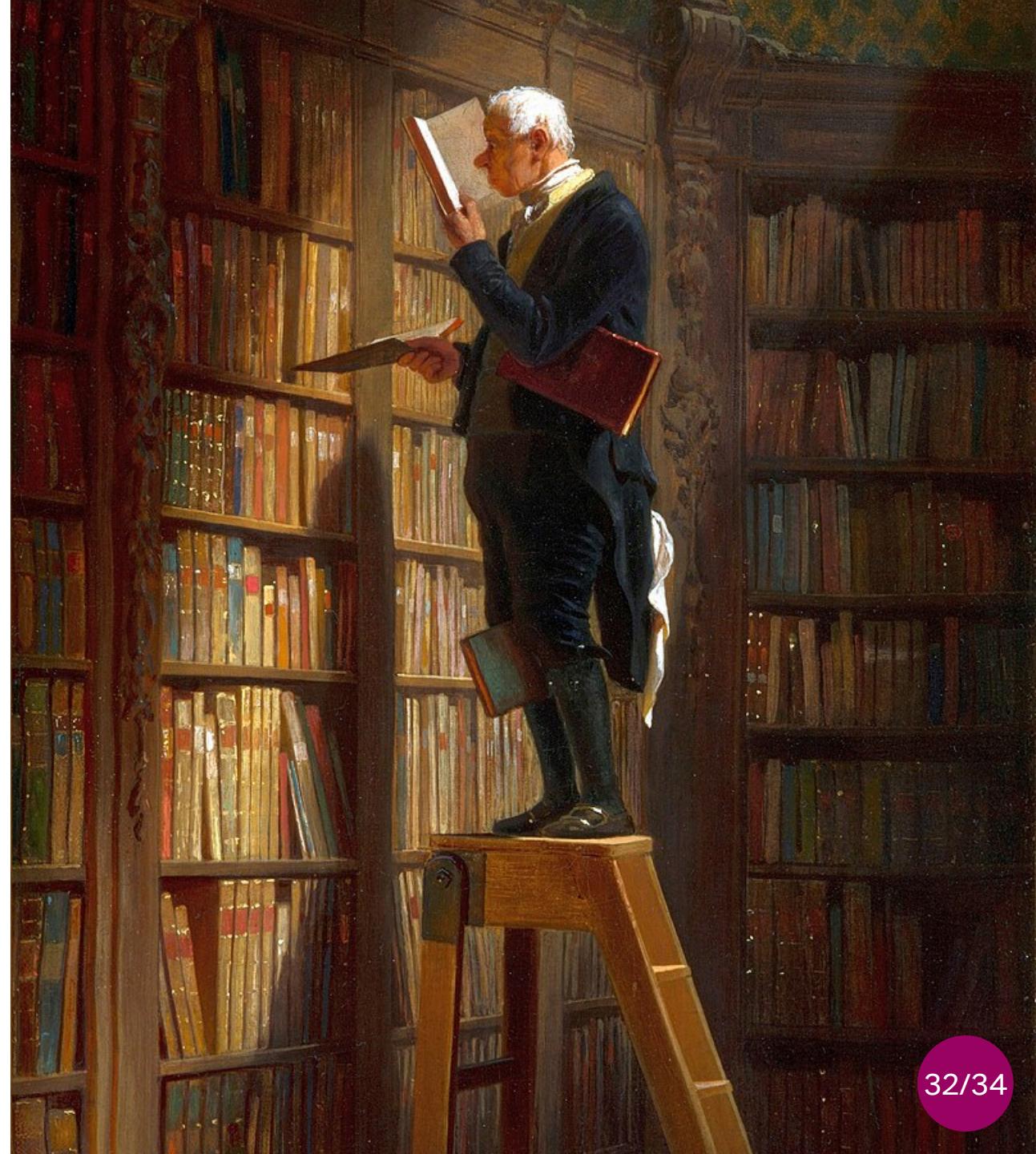
Descrizione

Dataset costituito da metadati descrittivi di pubblicazioni scientifiche che trattano di *biblioteconomia*.

Opportunità

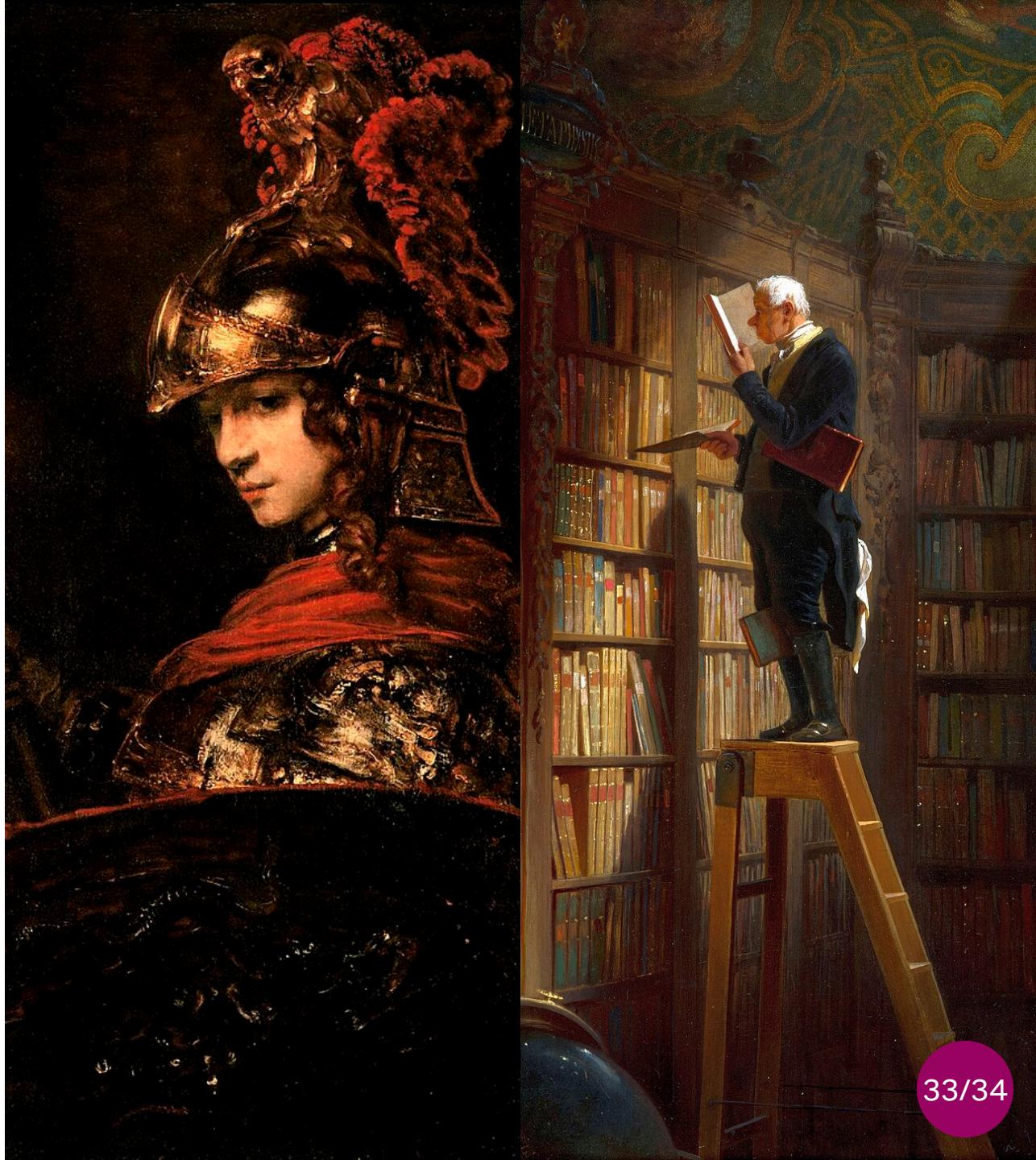
Implementazione di un progetto di **gestione dei dati** che coinvolge:

- 1 o 2 domande di ricerca;
- creazione di un database;
- un' **analisi bibliometrica** sulla base delle domande scelte;
- presentazione del processo e risultati.



Ne ripareremo, ma intanto questa è l'idea generale

- Scelta della modalità (in gruppo o individuale);
- Scelta di uno dei due dataset;
- Modellazione, pulizia e visualizzazione dei dati tramite **SQL** (e, in particolare, **DB Browser for SQLite**);
- Documentazione non strutturata con **Markdown**;
- Documentazione strutturata con **standard di metadati**;
- Pubblicazione dei materiali su **GitHub** e **Zenodo**;
- Presentazione orale.



Digital Humanities e Data Management per i Beni Culturali (2025/2026)

01. Origini

- Mail: sebastian.barzaghi2@unibo.it
- ORCID: [0000-0002-0799-1527](https://orcid.org/0000-0002-0799-1527)
- Sito: sebastian.barzaghi2

