

# Digital Humanities

presented by

Vandana Jha



# Agenda

1. What is Digital Humanities?
2. Historical Evolution
3. Core Workflows & Methods
4. Tools & Platforms
5. Projects & Case Studies
6. Future & Critical Perspectives
7. Resources

# Agenda

1. What is Digital Humanities?
2. Historical Evolution
3. Core Workflows & Methods
4. Tools & Platforms
5. Projects & Case Studies
6. Future & Critical Perspectives
7. Resources

# What is Digital Humanities - Definition

- “The intersection of humanities and digital technologies” — *Drucker & Terras*
- Combines humanities disciplines with computational tools
- Supports analysis, visualization, preservation, and public engagement
- An interdisciplinary field combining humanities with digital technologies. Encompasses research, teaching, and publishing.



# What is Digital Humanities - Definition

- Digital technologies are electronic tools, systems, devices and resources that generate, store or process data:
  - Application softwares (tools):
    - Text editors, text analysis, photo editing, text encoding, video editing, social media, ...
  - Processes (actions):
    - Document scanning, OCR, encoding, indexing, referencing, Data formatting, Data analysis, Data converting, ...
  - Digital data (contents):
    - Digital text, digital audio, digital images, digital video, web pages and websites, data and databases, electronic books...

# What is Digital Humanities - Core Components

- Text encoding
- Data mining
- GIS & spatial mapping
- Network analysis
- Digital publishing & archiving



# What is Digital Humanities - Value Proposition

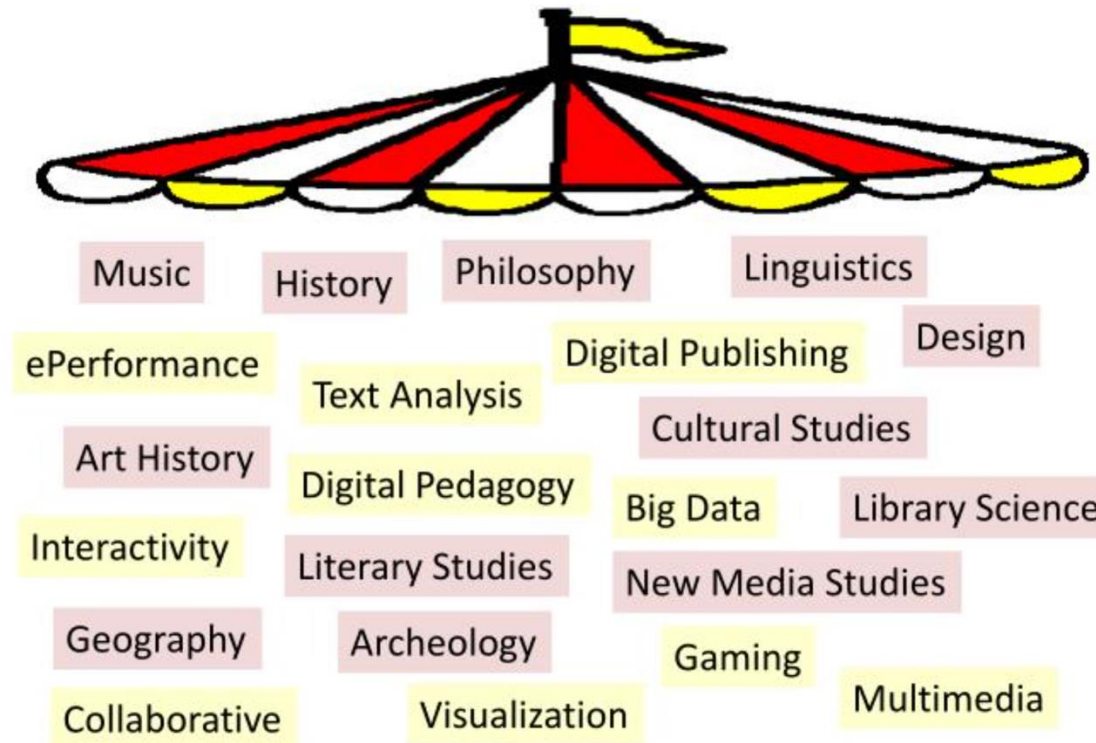
- Scalability for large corpora
- Enhances accessibility
- Fosters interdisciplinary critique
- Public engagement and co-creation

# What is Digital Humanities - Interdisciplinary Reach

Fields involved:

- History
- Linguistics
- Media Studies
- Computer Science
- Art & Design

‘Big Tent’ Digital Humanities





# Agenda

1. What is Digital Humanities?
- 2. Historical Evolution**
3. Core Workflows & Methods
4. Tools & Platforms
5. Projects & Case Studies
6. Future & Critical Perspectives
7. Resources

# Historical Evolution

- **Early Beginnings (1946–49)**

- In 1949, with the help of IBM, Roberto Busa, an Italian priest, embarks on a major computer-assisted edition project. He was the author of the Index Thomisticus, a complete lemmatization of the works of Saint Thomas Aquinas and of a few related authors.
- Seen as the first digital humanities project

- **Humanities Computing Era (1960–80s)**

- Early text corpora & computing centers
- Formation of ALLC (**Association for Literary and Linguistic Computing**) which later becomes the EADH in 1973

# Historical Evolution

- **TEI & Markup Standards**

- TEI (Text Encoding Initiative) launched in 1987
- Guidelines published in 1994 for consistent encoding.

```
<?xml version="1.0" encoding="UTF-8" ?>
<teiHeader xml:lang="en">
  <fileDesc>
    <titleStmt>
      <title level="s">r/OpenChristian</title>
      <title level="a">7 ways the new PC(USA) Hymnal is queer.</title>
      <ns1:include href="https://276ddc12-4e43-4e64-94ba-fab7481cb713.ka.bw-cloud-instance.org/base-repo/api/v1/dataresources/31ba993c-8907-4c13-a996-9f275bc46088/data/ingest_reddit_respStmt.xml"/>
    </titleStmt>
    <ns1:include href="https://276ddc12-4e43-4e64-94ba-fab7481cb713.ka.bw-cloud-instance.org/base-repo/api/v1/dataresources/f8274559-aaa3-420c-97c8-05704817159a/data/ingest_publicationStmt.xml"/>
    <ns1:include href="https://276ddc12-4e43-4e64-94ba-fab7481cb713.ka.bw-cloud-instance.org/base-repo/api/v1/dataresources/f514e54e-b6b3-4337-add2-7533cfc4b855/data/ingest_reddit_sourceDesc.xml"/>
  </fileDesc>
  <ns1:include href="https://276ddc12-4e43-4e64-94ba-fab7481cb713.ka.bw-cloud-instance.org/base-repo/api/v1/dataresources/349107ed-95e3-40ba-915e-313d84fab5a5/data/ingest_redditOpenChristian_encodingDesc.xml"/>
  <profileDesc>
    <creation>
      <date type="thread" from="2014-01-01T20:36:54" to="2014-01-01T20:36:54">from 01 Jan 2014 till 01 Jan 2014</date>
      <date type="scrape" when="2022-06-10T04:55:55.550961">10 Jun 2022</date>
      <date type="file" when="2023-07-04T10:58:15.772424">04 Jul 2023</date>
      <date type="tokenization" when="2023-07-04T12:55:42.846169">04 Jul 2023</date>
    </creation>
  </profileDesc>
</teiHeader>
```

```
<?xml version="1.0" encoding="UTF-8" ?>
<text xml:lang="sa-Latn">
  <body>
    <div type="book" n="Rāmāyaṇa" xml:id="b.143">
      <div type="kāṇḍa" n="Rām, Bā">
        <div type="chapter" n="1" xml:id="c.1067">
          <lg>
            <l xml:id="l.110534">
              <choice n="sandhi">
                <orig>
                  <w xml:id="w.318668_318669_318670.s">
                    <w>tapasvādhyaṇiratam</w>
                  </w>
                </orig>
                <reg>
                  <w xml:id="w.318668">tapas</w>
                  <w xml:id="w.318669">svādhyā</w>
                  <w xml:id="w.318670">niratam</w>
                </reg>
              </choice>
            </l>
          </lg>
        </div>
      </div>
    </div>
  </body>
</text>
```

# Historical Evolution

- **Rebranding & Growth (2004–2006)**

- “Digital Humanities” replaces “Humanities Computing”
- NEH (National Endowment for the Humanities) Office of Digital Humanities established
- Increased funding & academic recognition

- **DH as a Field**

- Rise of DH centers (Stanford, King’s College, etc.)
- THATCamps, conferences, journals (e.g., *Digital Humanities Quarterly*)

# Agenda

1. What is Digital Humanities?
2. Historical Evolution
- 3. Core Workflows & Methods**
4. Tools & Platforms
5. Projects & Case Studies
6. Future & Critical Perspectives
7. Resources

# Core Workflows & Methods

- **Text Encoding (TEI/XML)**

- Enables structured markup of literary/historical texts
- Used in scholarly digital editions

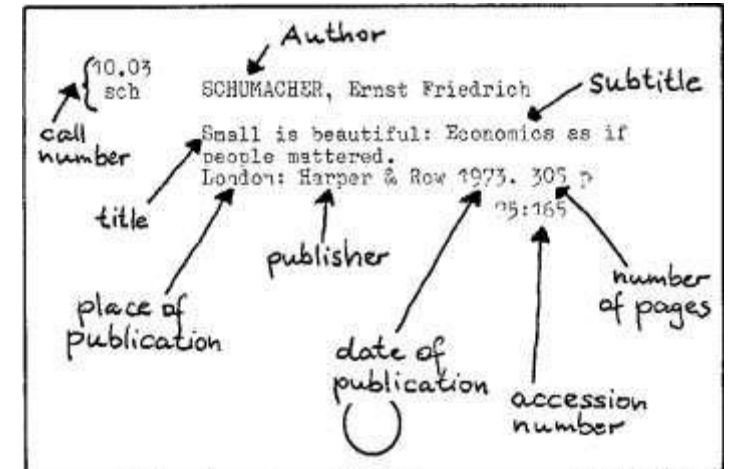


- **Text Mining & Corpus Linguistics**

- Frequency analysis, concordance, collocations
- Tools: NLTK, AntConc

- **Topic Modeling & Sentiment Analysis**

- Discover latent themes (e.g., LDA)
- Track sentiment trends across corpora



# Core Workflows & Methods

- **Network Analysis**

- Visualize relationships among authors, locations, or concepts
- Tools: Gephi, Cytoscape

- **Spatial Humanities & GIS**

- Maps historical data using QGIS, ORBIS
- Reconstructs past environments

- **Digital Editions & Archives**

- Creating scholarly editions of texts
- Preserves context, layers of meaning

# Digital Editions

**The Walt Whitman Archive**  
Ed Folsom & Kenneth M. Price, Editors

SUPPORT THE ARCHIVE

Search the Archive

**Published Works**  
[Books by Whitman](#); [Periodicals](#); [Translations](#)

**In Whitman's Hand**  
[Manuscripts](#); [Notebooks](#); [Scribal Documents](#)

**Life & Letters**  
[Biography](#); [Chronology](#); [Letters](#)

**Commentary**  
[Bibliography](#); [Criticism](#); [Disciples](#); [Reviews](#)

**Pictures & Sound**  
[Audio](#); [Portraits of Whitman](#)

**Resources**  
[Downloads](#); [Finding Aids](#); [Search](#); [Teaching](#); [Tools](#)




**News & Updates**


August 2012  
With the support of the Obermann Center at the University of Iowa, we've revised and expanded the [Translations](#) section of the Archive, which now includes more than thirty versions of Whitman's "Poets to Come" in five languages, as well as original translations of "Chants Democratic 14" from the 1860 *Leaves*. Read more about these and other changes to the site [here](#).

June 2012  
All volumes of Horace Traubel's [With Walt Whitman in Camden](#) are

# Digital Archives

*The Proceedings of the* **OLD BAILEY**  *London's Central Criminal Court, 1674 to 1913*

[Home](#) | [Search](#) | [About The Proceedings](#) | [Historical Background](#) | [API](#) | [The Project](#) | [Contact](#)



**Home Page**  
[Search](#)  
[About the Proceedings](#)  
[Historical Background](#)  
[API](#)  
[The Project](#)  
[Copyright & Citation Guide](#)  
[Contact](#)  
[Research and Study Guides](#)  
[Login / Register](#)

**The Proceedings of the Old Bailey, 1674-1913**

A fully searchable edition of the largest body of texts detailing the lives of non-elite people ever published, containing 197,745 criminal trials held at London's central criminal court. If you are new to this site, you may find the [Getting Started](#) and [Guide to Searching](#) videos and tutorials helpful.

To search the Proceedings use the boxes on the right or go to the [Search Pages](#).

**SEARCH**  
*the Proceedings*

Keyword(s)

Reference No.

Search In

[More Search Options](#)

ON THIS DAY IN **1784**



# Core Workflows & Methods

- **Crowdsourcing & Engagement**

- Public participation in transcription, annotation
- Examples: Transcribe Bentham, Zooniverse

- **Metadata & Preservation**

- Structured metadata ensures discoverability
- Challenges: format obsolescence, long-term storage

- **Reflexivity in DH**

- Critical approaches to data practices
- Attention to transparency, algorithmic bias

# Core Workflows & Methods

- **Open Access & Public Scholarship**
  - Removing paywalls
  - Enabling remixing and reuse of cultural data
- **Ethical Concerns in DH**
  - Surveillance, digital colonialism, algorithmic harm
  - Emphasis on equity, access, and inclusion

# Agenda

1. What is Digital Humanities?
2. Historical Evolution
3. Core Workflows & Methods
- 4. Tools & Platforms**
5. Projects & Case Studies
6. Future & Critical Perspectives
7. Resources

# Tools & Platforms

- **TEI/XML**

- XML-based standard for rich text markup
- Supported by many archives

- **Voyant Tools**

- Web-based suite for text analysis
- Visualizations: word clouds, trend graphs

- **Python & R**

- Libraries: NLTK, spaCy, Gensim (Python); Tidytext, Quanteda (R)
- Supports reproducible workflows



Voyant tool: <https://voyant-tools.org/?lang=en>



Add Texts 🔍 ?

Type in one or more URLs on separate lines or paste in a full text.

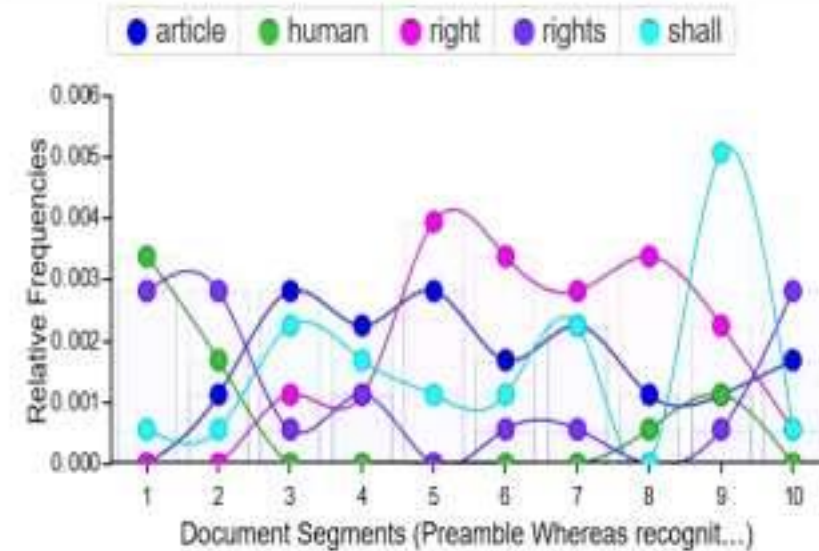
 Open  Upload  Reveal



## Preamble Whereas recognition of the inherent di...

### Preamble

Whereas recognition of the inherent dignity and of the equal and inalienable rights of all members of the human family is the foundation of freedom, justice and peace in the world

Terms: 
 ?

 ?

Reset

Display

This corpus has 1 document with 1,779 total words and 534 unique word forms. Created 2 seconds ago.

Vocabulary Density: 0.300

Average Words Per Sentence: 30.2

Most frequent words in the corpus: right (33); article (30); shall (27); rights (21); human (12)

Items: 
 ?

33 context

expand

# Voyant tool: <https://voyant-tools.org/?lang=en>

- The 3 main parts of the result screen are:
  - On the left, the area that shows the clouds Tags (Cirrus), the occurrences of words in the text (Terms) and semantic links between terms (Links)
  - The middle area displays the text that is the subject of analysis. Select a word and it will be highlighted throughout the text years
  - The right area shows the curve of the most frequent words, the number of occurrences of words and their proportions relative to the entire text
- The lower half of the screen displays:
  - Left the summary of the document analyzed, the documents in the case of a body being analyzed, and phrases (group of words as a unit)
  - On the right, the context of each of the words extracted, and a line of bubbles of different sizes depending on the occurrence of each term.





# AntConc

A freeware corpus analysis toolkit for concordancing and text analysis

[AntConc Homepage](#)

[Screenshots](#)

[Help](#)

[License](#)

[Version History](#)

## Downloads

### Official releases

- [Windows \(Installer\) \(4.3.1\)](#)
- [Windows \(Portable\) \(4.3.1\)](#)
- [MacOS 10/11-Silicon \(4.3.1\)](#)
- [MacOS 10/11-Intel \(4.3.1\)](#)
- [Linux \(Portable\) \(4.3.1\)](#)

### AntConc 3x series

- [Windows \(3.5.9\)](#)
- [MacOS 10/11 \(3.5.9\)](#)
- [Linux \(portable\) \(3.5.9\)](#)

### Beta releases







# Tools & Platforms

- **GIS Tools**

- QGIS, ArcGIS for spatial analysis
- QGIS: <https://www.qgis.org/fr/site/>
- ArcGIS: <https://www.arcgis.com/index.html>
- Used for mapping literature, migration, battles

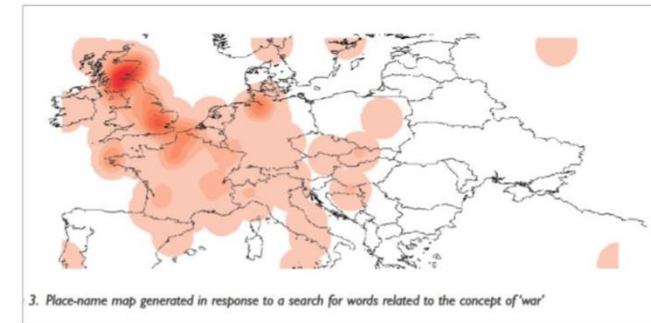
- **Network Tools**

- Gephi, Cytoscape for network graphs
- Import from CSV or GraphML

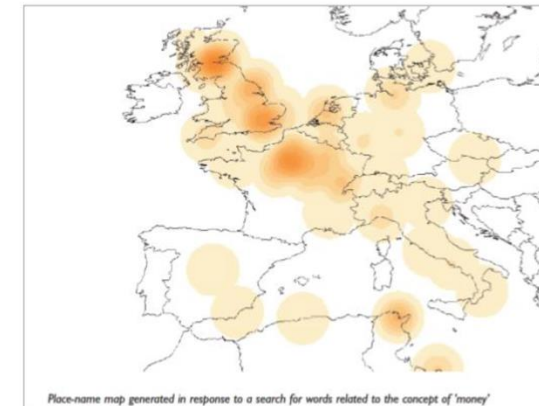
- **Visualization Tools**

- Tableau, Flourish, D3.js for interactive visuals

GIS Mapping: Topic 'War'



GIS Mapping: Topic 'Money'



# Tools & Platforms

- **Digital Publishing**

- Omeka (archives), Scalar (multimedia narratives), Pressbooks (open publishing)

- **Code & Collaboration Platforms**

- GitHub for versioning
- Jupyter for documentation and analysis
- Zooniverse for public contributions

# Agenda

1. What is Digital Humanities?
2. Historical Evolution
3. Core Workflows & Methods
4. Tools & Platforms
- 5. Projects & Case Studies**
6. Future & Critical Perspectives
7. Resources

# Projects & Case Studies

- **Rossetti Archive & WWP**
  - Digital editions with markup
  - Emphasis on literary and feminist scholarship
- **Project Gutenberg & Ngram Viewer**
  - Large-scale digital text access
  - Word frequency over centuries
- **Republic of Letters & ORBIS**
  - Mapping historical networks
  - Geospatial simulations of Roman travel

# Projects & Case Studies

- **CLiC Dickens & Corpus Linguistics**
  - KWIC, dialog modeling in literary corpora
  - Insights into character interaction
- **Crowdsourcing Projects**
  - Transcribe Bentham
  - Shakespeare's World
  - Old Weather (WWI ship logs)
- **Cultural Heritage Archives**
  - Europeana
  - Digital VOC
  - Perseus Digital Library

# Projects & Case Studies

- **Archival Digitization**

- Charles Darwin's Library
- British Museum manuscripts

- **Thematic Highlights**

- Multilingual DH: Parallel corpora, translation tech
- Multimodal DH: Images, sound, video, VR
- Feminist & Postcolonial DH: Emphasis on recovery and critique
- DH in the Global South: Infrastructure & access
- Community-Driven DH: Indigenous knowledge, oral history

# Agenda

1. What is Digital Humanities?
2. Historical Evolution
3. Core Workflows & Methods
4. Tools & Platforms
5. Projects & Case Studies
- 6. Future & Critical Perspectives**
7. Resources



# Future & Critical Perspectives

- **Challenges Ahead**

- Data bias
- Long-term digital preservation
- Training gaps and tool accessibility

- **Opportunities**

- Cultural analytics
- AI-enhanced interpretation
- Virtual heritage, AR, storytelling

# Agenda

1. What is Digital Humanities?
2. Historical Evolution
3. Core Workflows & Methods
4. Tools & Platforms
5. Projects & Case Studies
6. Future & Critical Perspectives
- 7. Resources**

# Resources

## • Further Reading & Communities

- *Digital Humanities Quarterly*
- DHNow
- ADHO
- DARIAH
- THATCamp



# Q&A and Feedback

Dr. Vandana Jha  
Data Exploitation Methods (DEM),  
Scientific Computing Center (SCC),  
Karlsruhe Institute of Technology (KIT)

<http://www.scc.kit.edu>