1 Digital humanities overview

This chapter covers an overview of digital humanities, beginning with a definition and an introduction to the components and workflows involved in digital projects. It details fundamental activities such as remediation, datafication, processing, presentation, and sustainability, explaining how each aspect shapes digital humanities research. This chapter also emphasizes the importance of understanding the ethical and critical implications of using digital tools and methods in humanistic work. Additionally, it outlines practical steps and considerations for structuring digital humanities projects and highlights key debates about the integration of technological methods with humanistic values.

1a What is digital humanities?

Exercise 1.1: Analyzing a project

Look at one example of a digital humanities project and see if you can describe it in terms of materials, processing, presentation, and participation.

How-to examples

- The Green Book Map project (https://publicdomain.nypl.org/greenbook-map/)
    - The Green Book Map project by the New York Public Library is a fascinating digital humanities initiative that documents and visualizes the locations listed in The Negro Motorist Green Book, a mid-20th-century travel guide that provided African American travelers with information on safe places to stay, eat, and visit during the era of segregation in the United States. Here's how it can be described in terms of materials, processing, presentation, and participation:
        - Materials:
            - **Primary sources**: The core material of the project is *The Negro Motorist Green Book*, which was published annually from 1936 to 1966. This guide listed hotels, restaurants, gas stations, and other businesses across the United States that were safe for African American travelers to visit. The project uses the content from these books, which has been digitized and geocoded.
            - **Geospatial data**: The project includes geospatial data corresponding to the addresses of the locations listed in *The Green Book*. The locations are plotted on an interactive map to give a spatial understanding of the distribution of these safe places.
            - **Historical context**: Alongside the raw data of locations, the project incorporates rich historical information, including images, context about the locations, and the broader social and political backdrop of segregation and travel for Black Americans.
        - Processing:
            - **Digitization and optical character recognition (OCR)**: The content from *The Green Book* has been digitized, likely using OCR to extract text from the printed books. This text is then processed and cleaned to ensure accuracy, as well as geocoded to match addresses to physical locations on a map.
            - **Geocoding**: The addresses listed in the original guide have been converted into geographic coordinates. This involves processing historical addresses, which may no longer exist or may have changed over time, and accurately placing them on a modern map.
            - **Data structuring**: The project has structured the listings into a database with fields like business type, location, and year of inclusion in *The Green Book*. This structured data allows for efficient querying and display on the interactive map.

- Presentation:
    - **Interactive map interface**: The Green Book Map is presented through an interactive web-based map. Users can explore locations by state or city and filter the map based on categories like hotels, restaurants, or gas stations. Each point on the map links to more detailed information about that location, including its address, the year it was listed, and sometimes images or other historical context.
    - **Timeline**: The map allows users to filter data by year, showing how the listings changed over time as *The Green Book* expanded. This gives a sense of historical progression and offers insight into how racial segregation impacted travel over different decades.
    - **User experience**: The project is designed for easy exploration, with intuitive map navigation and a simple layout that allows users to see both the scale and details of the listed locations. The combination of geographic and historical context makes it a visually compelling and informative presentation of *The Green Book*'s contents.
- Participation:
    - **User interaction**: The project allows for deep user interaction. Visitors can explore the data by navigating through the map, clicking on specific points to learn more about locations, and using filters to customize their experience. This interactive element encourages users to engage with the historical data in a meaningful way.
    - **Public access and education**: The project opens a window into a critical part of African American history, making it accessible to a broad audience, from scholars to the general public. The public nature of the platform invites users to learn and reflect on the historical context of segregation and travel in America.
    - **Community engagement potential**: Although the project does not currently have a feature for user submissions or crowd-sourced data, there is potential for future participation in the form of community contributions, where users might add information about particular locations or submit personal stories related to traveling with *The Green Book*.
- Summary: The Green Book Map is a well-executed digital humanities project that takes primary source material from *The Negro Motorist Green Book*, processes it into geospatial data, and presents it through an accessible and educational interactive map interface. The project invites user participation by allowing exploration of historical data in a dynamic way, making it both a tool for historical research and public education.

- Bitter Aloe Project (https://bitteraloeproject.createuky.net/)
    - Project Review by Nabeel Siddiqui (https://reviewsindh.pubpub.org/pub/bitter-aloe-project/release/1)
- South Asian Canadian Digital Archive Project (https://sacda.ca/index.php)
    - Project Review by Luis Meneses (https://reviewsindh.pubpub.org/pub/south-asian-canadian-digital-archive/release/4)
- Digital Pasifik (https://digitalpasifik.org/)
    - Project Review by Kate Topham (https://reviewsindh.pubpub.org/pub/pacific-virtual-museum/release/3)
- Native Land Digital (https://native-land.ca/)
    - Project Review by Thiago da Costa Oliveira (https://reviewsindh.pubpub.org/pub/native-land-digital/release/1)
- Mapping Black California (https://mappingblackca.com/)
    - Project Review by Ashanté Reese (https://reviewsindh.pubpub.org/pub/mapping-black-california/release/1)

Try it for yourself with the following projects:

- Historical photographs of China (http://www.hpcbristol.sjtu.edu.cn/)
    - Consider bilingual or Chinese-language DH projects: https://www.njlit.com or https://dhc.library.sh.cn/. How might your approach to sources, process, and presentation change for multilingual digital projects? What else might you need to consider when working with languages that are read top to bottom or right to left?
- In the Spotlight (http://playbills.libcrowds.com/)
- Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC) (https://catalog.paradisec.org.au/)
- Renaissance Lyon: Le Plan Scénographique c. 1550 (http://www.renlyon.org)
- Arqueología histórica de Santiago (https://instagram.com/arqueologiahistoricasantiago)

Recommended readings

Croxall, Brian. 2018. "Taxonomy of Research Topics in Digital Humanities." https://github.com/dhtaxonomy/TaDiRAH/tree/master.

Flanders, Julia, and Fotis Jannidis, eds. 2019. *The Shape of Data in the Digital Humanities: Modeling Texts and Text-Based Resources*. London: Routledge. https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=1926341.

Greenspan, Brian. 2019. "The Scandal of Digital Humanities." In *Debates in Digital Humanities*, edited by Lauren Klein and Matthew K. Gold, 92–95. Minneapolis: University of Minnesota Press. https://www.jstor.org/stable/10.5749/j.ctvg251hk.12?seq=1.

Liu, Alan. 2012. "The State of the Digital Humanities: A Report and a Critique." *Arts and Humanities in Higher Education* 11 (1–2): 8–41. https://doi.org/10.1177/1474022211427364.

Viola, Lorella. 2023. *Humanities in the Digital: Beyond Critical Digital Humanities*. Cham: Palgrave Macmillan. https://doi.org/10.1007/978-3-031-16950-2.

1b Principles and scenarios for digital humanities

Exercise 1.2: Basic components

Using one of the scenarios below, describe the components of the project (materials-processing-presentation) to see if you understand the concepts involved. (See the textbook for full descriptions.)

- Scholar A is studying the history of collecting Indigenous artifacts in a major natural history museum. The research question is how catalog records, descriptions, acquisition records, field notes, and other documentation imposed colonial views. This is a text analysis and topic modeling project.
- Scholar B is examining the role of geography on a Caribbean island nation in shaping the outcome of a slave rebellion and its relation to current development plans for tourism. This is a mapping project, but the available source materials are not evident, and it may need to evolve into a 3D modeling project with AR immersive features.
- Scholar C is focusing research on the shaping of queer identities in private recordings of performance art of the 1960s in New York City and their influence on popular perception of gay culture when they became available online in the 2000s. This is an audio-visual project but also involves topic modeling, social network analysis, text analysis, and data mining.
- Scholar D is interested in creating a virtual reconstruction of a Hindu festival site using a collection of three thousand photographs, field notes, drawings, and other materials taken before and after the site was badly damaged in a recent conflict. This is a 3D modeling project.

How-to example

Based on the first scenario, the project Scholar A is working on involves multiple components that span materials, processing, and presentation stages. Here's a breakdown of each:

Materials:

- **Primary sources**: Scholar A will be working with catalog records, descriptions, acquisition records, field notes, and other documentation related to Indigenous artifacts in a major natural history museum. The focus is on how these materials imposed colonial views during the time of collection.
- **Digital sources**: It's important to determine whether these documents have been digitized, and if so, what formats are available (e.g., PDFs, text files). If they are not digitized, Scholar A must consider the appropriate format for digitization, such as text files for analysis or structured formats like CSV for data organization.
- **Acquisition records**: Research will include understanding the provenance of the artifacts and identifying who has the rights to these cultural properties. Cultural sensitivity is critical here, especially in engaging with Indigenous communities for ethical practices in managing these materials.

Processing:

- **Data preparation**: The project begins with gathering, reviewing, and digitizing (if necessary) the catalog records and related documentation. Scholar A will also assess the scale and timeframe for this study, determining how many records and artifacts will be included.
    - **Standardization**: An initial study of classification vocabulary terms will be undertaken to determine how standardized the terms have been over time, and how these terms may reflect colonial bias.
- **Cultural consultation**: Engaging with Indigenous communities to address issues of cultural property rights and sensitivities around the data is an important part of this phase. These communities may provide critical insights into how these records should be handled, named, and displayed.

- **Text analysis and topic modeling**: Scholar A will use OCR software to convert physical or image-based documents into text, if necessary. The resulting text data will be analyzed using text analysis tools to identify key terms, and a topic model will be created to explore themes of colonialism and Indigeneity.
    - **Control group**: For comparative purposes, a control set of non-Indigenous artifact records from the same period will be examined to identify differences in classification and description approaches.

Presentation:

- **Visualization**: Once the data has been processed and analyzed, Scholar A will create visualizations to display the frequency of terms associated with colonialism and Indigeneity over time. These visualizations will help demonstrate the evolving or persistent biases in the museum's records.
- **Reparative work and dialogue**: Based on the findings, Scholar A will discuss potential reparative actions with museum stakeholders and Indigenous communities. This might include reclassification, renaming, or the ethical display and handling of artifacts.
- **Ethical access and dissemination**: The final stage of the project involves creating a digital repository or interface that ensures ethical access to the cataloged records. The digital platform will need to consider contemporary practices for culturally sensitive data sharing, especially when dealing with artifacts that are still subject to tribal or Indigenous laws.

Overall tools and technologies:

- **OCR software** (e.g., Adobe Acrobat) to digitize documents.
- **Text analysis platforms** (e.g., Python's Natural Language Toolkit [NLTK] or topic modeling tools like Mallet) to analyze the digitized records and extract terms and patterns.
- **Visualization tools** (e.g., Tableau or Flourish) to visually represent the results, such as the frequency of colonial terms across time periods.
- **Structured data formats** (e.g., spreadsheets, databases using standards such as Categories for the Description of Works of Art [CDWA (https://www.getty.edu/research/publications/electronic_publications/cdwa/)] or Dublin Core (https://www.dublincore.org/)) for organizing the data in ways that align with cultural cataloging standards.

In summary, Scholar A's project involves the collection and ethical management of museum records related to Indigenous artifacts. It focuses on understanding how these records reflect colonial viewpoints and works to reframe them through consultation with Indigenous communities, using digital tools for text analysis, topic modeling, and visualization. Each phase of the project is designed to ensure both historical understanding and ethical digital access.

How-to example—extra credit

Work through this example in the same way as the example above.

Scholar X is assessing the ways characterizations of gender appeared in posters and broadsheets with woodcuts during the Mexican Revolution. How are events portrayed and do gender and ethnicity play a role? How is gender portrayed visually? How do the terms of moral judgment change over time?

This is both a text and an image project. The corpus of materials is clear. The work will involve text analysis and data mining to produce visualizations.

Workflow

- Identify which ballads and broadsheets to use and where they are located.
- Make a spreadsheet in which events are identified with gendered and racially classified individuals.
- Determine how gender and race are being defined.
- Determine how moral judgments are defined and expressed.
- Determine which terms of judgment are identified with race and gender.
- Decide whether you will rely on tags/metadata descriptions or extract the texts from the printed materials.
- Determine which features of the images are significant (clothing, accessories, settings, facial expressions, body language, gestures, etc.).
- Create a vocabulary for tagging posters and broadsheets with metadata.
- Create alt.image texts to describe the images for visually impaired users. Consider the issues in describing potentially offensive material.
- Enter the data into a spreadsheet.
- Generate visualizations that chart the use of gendered and racist terms.
- Determine how to chart moral judgments linked to these categories in a meaningful way.

Discussion

The challenges here are the amount of labor involved in creating a meaningful tag set, applying it consistently to gender and racial identity and their connection to moral judgments, and determining a visualization that allows multiple variables to be displayed. Decide if the final project should be static images or a dataset to which filters and queries can be applied. If the latter, then the controlled vocabulary should be offered to the viewer. If the full text of the posters and broadsheets is going to be made accessible, they will have to be remediated (automatically, with correction by hand or keyboarded). The intellectual decisions by which moral judgments are modeled should be documented.

Exercise 1.3: Workflow

Using another of the scenarios, describe the ways the activities of mediation, datafication, processing/analytics, presentation, and sustainability need to be addressed. Assess how your design will meet accessibility standards.

How-to examples

Using the first scenario again, answers may look like the following:

MEDIATION

Mediation in this project begins with determining the availability and format of source materials. Scholar A must assess whether the catalog records, descriptions, field notes, and acquisition records are digitized, and if not, decide on the appropriate format for digitization (e.g., PDFs, OCR-readable text, or images). In addition, mediation involves engaging with the Indigenous communities whose cultural property is being analyzed. Since these materials may hold significant cultural value and sensitivity, it is essential to consult with these communities to ensure ethical handling, classification, and potential return of artifacts. Scholar A must also mediate between different stakeholders, such as museum curators, archivists, and Indigenous representatives, to facilitate transparent dialogue regarding the ownership and use of these artifacts.

DATAFICATION

Once the materials are collected and digitized, Scholar A must decide how to structure and encode the data. Catalog records, acquisition documents, and field notes will likely contain unstructured text that needs to be converted into structured formats for analysis. Decisions need to be made about whether to store the data in spreadsheets with specific fields (e.g., object name, classification, date, origin) or whether to retain unstructured text for deeper textual analysis. Scholar A will also develop a controlled vocabulary to classify terms, particularly those related to gender, race, and colonialism, ensuring consistency in how terms are tagged across the dataset. This step is crucial, as it allows for accurate text mining and topic modeling in later stages. Special attention must be given to how colonial terms were used historically in these documents and how to encode such terms without reinforcing colonial biases.

PROCESSING/ANALYTICS

In the analytics phase, Scholar A will run the digitized texts through text analysis programs, identifying key terms, phrases, and patterns that reflect colonial views. This may include measuring the frequency of terms associated with Indigenous objects, examining how their descriptions shifted over time, and identifying patterns of classification based on colonial perspectives. To ensure the results are accurate, a control set of non-Indigenous artifacts from the same periods should be used for comparison. Text mining and topic modeling tools like Python's NLTK or other machine learning platforms will help extract themes related to colonialism and Indigeneity. Scholar A will then analyze how standardized classification terms were used across time and whether shifts in the descriptions of Indigenous objects coincide with larger colonial narratives or political changes.

PRESENTATION

Scholar A will need to decide how to present the findings in a way that is meaningful, transparent, and engaging for both academic and public audiences. Visualization tools such as Tableau or Flourish could be used to create dynamic charts or graphs showing the frequency of colonial terms over time or visualizing how classifications of Indigenous objects changed across specific periods. Scholar A might also create topic models that display key themes related to colonialism and the representation of Indigenous cultures, offering a visual map of the relationships between terms and classifications. Additionally, Scholar A should present findings in ways that respect the input and feedback from Indigenous communities. This could include creating public-facing reports or digital exhibitions that highlight reparative work or shifts in the narrative around these artifacts.

SUSTAINABILITY

Sustainability in this project will involve creating a plan to ensure that the digitized records, the dataset, and visualizations can be preserved for future research. Scholar A should consider developing the dataset using metadata standards like the CDWA or Dublin Core and output data in file formats such as CSV or XML to ensure interoperability and long-term use. Further, Scholar A should document all intellectual decisions made during the project, such as how colonial terms were categorized and how reparative work was conducted. This transparency is critical for future scholars who may build on this work or apply similar methodologies to other collections. Scholar A must also ensure that the project aligns with the long-term goals of the Indigenous communities involved, considering ways that the findings might inform future restitution, reinterpretation, or repatriation efforts.

ACCESSIBILITY

As a text analysis project, document remediation will be an important step to ensure all research and outcomes is accessible to people with disabilities, particularly those using assistive technology like screen readers. In addition, Scholar A may opt for a platform such as *Mukurtu* (https://mukurtu.org/), which is designed with Indigenous communities in mind. Levels of access can be set so that certain sensitive information is not visible to all users, allowing communities to control what information they choose to share.

In summary, Scholar A's project will involve digitizing and structuring large datasets, running complex text and image analyses, presenting findings through thoughtful visualizations, and ensuring that the project remains sustainable and ethically responsible, particularly with respect to the Indigenous communities whose cultural heritage is at stake.

Exercise 1.4: Ethical concerns

In which scenarios shared in Exercise 1.2 are issues of privacy most likely to need consideration? Cultural appropriation and ownership? Intellectual property?

How-to example

PRIVACY CONCERNS

Scholar C: Privacy is a significant concern here, as the research involves private recordings of performance art from the 1960s in New York City. Since many of these performances were private or intended for closed audiences, Scholar C will need to address intellectual privacy, ensuring that the rights of the performers and participants are respected. There may be issues of consent, as performers from the 1960s may not have expected these recordings to be widely disseminated or studied. In addition, determining liability regarding the use of these private materials—especially when they later became public—is a critical issue.

CULTURAL APPROPRIATION AND OWNERSHIP

Scholar A: The primary concern for Scholar A is cultural appropriation and ownership. Since the project deals with Indigenous artifacts, Scholar A must carefully address how these objects were acquired and whether their current custodianship in a natural history museum is ethically sound. The documentation of these objects may impose colonial views, and reparative work with Indigenous communities is necessary to ensure cultural ownership is acknowledged. Consultation with these communities about how their cultural property is cataloged and described is critical, as naming conventions and classification systems may perpetuate colonial biases. Furthermore, the project will need to navigate legal and cultural frameworks around the ownership and potential restitution of these artifacts.

INTELLECTUAL PROPERTY

Scholar C: Intellectual property is a key issue in this scenario as well. Scholar C needs to determine who owns the rights to the performance art recordings and obtain appropriate releases for research use. Since these tapes capture performances, music, and potentially copyrighted work, navigating the intellectual property rights of the performers and creators will be crucial. Without proper releases or agreements, Scholar C may face legal challenges when using or disseminating the materials. Furthermore, Scholar C will need to create metadata for describing the performances, which involves intellectual decisions that must be credited appropriately.

OTHER ISSUES

Scholar B: While privacy and intellectual property are less prominent issues in Scholar B's project, cultural appropriation and ownership might emerge, especially regarding how the perspectives of enslaved people and colonial forces are

represented. Scholar B must carefully construct an ethical and accurate portrayal of these historical events while considering the asymmetry of available documentation between the two populations.

Scholar D: Intellectual property and ownership are secondary concerns in Scholar D's work but still relevant, especially regarding the materials created by community groups, earlier anthropologists, or scholars. These potentially raise issues of cultural ownership and appropriation. However, the focus is more on ensuring accurate reconstruction rather than privacy or intellectual property per se. Members of the Hindu community should be actively consulted and involved in the project design.

IN SUMMARY

- Privacy: Scholar C's project has the most pressing privacy concerns, due to the private nature of the performance recordings and their later availability to the public.
- Cultural appropriation and ownership: Scholar A's project is most focused on issues of cultural appropriation and ownership, given the colonial history of the Indigenous artifacts.
- Intellectual property: Scholar C's project also faces significant intellectual property issues due to the ownership and use of private recordings of the performances, photographs, and other legacy documents.

In all of these cases, careful consideration of these issues is necessary to ensure ethical and legal compliance in the research process.

Recommended readings

Gold, Mathew, and Lauren F. Klein. 2023. "Introduction: The Digital Humanities, Moment to Moment." In *Debates in the Digital Humanities*. Minneapolis: University of Minnesota Press. https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=3406850.

Posner, Miriam. 2013. "How Did They Make That?" https://miriamposner.com/blog/how-did-they-make-that/.

Risam, Roopika. 2015. "Beyond the Margins: Intersectionality and the Digital Humanities." *Digital Humanities Quarterly* 9 (2). www.digitalhumanities.org/dhq/vol/9/2/000208/000208.html.

Bibliography

Barats, Christine, Schafer, Valerie, and Fickers, Andreas. 2020. "Fading Away… The Challenge of Sustainability in Digital Studies." *Digital Humanities Quarterly* 14 (3). https://www.digitalhumanities.org/dhq/vol/14/3/000484/000484.html.

Blanchette, Jean-François. 2011. "A Material History of Bits." *Journal of American Society for Information Sciences and Technology* 62 (6). https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.21542.

Buolamwini, Joy. 2023. *Unmasking AI: My Mission to Protect What Is Human in a World of Machines*. First edition. New York: Random House.

Costanza-Chock, Sasha. 2020. *Design Justice: Community-Led Practices to Build the Worlds We Need.* Cambridge: The MIT Press. https://doi.org/10.7551/mitpress/12255.001.0001.

Criado-Perez, Caroline. 2019. *Invisible Women: Data Bias in a World Designed for Men*. New York: Abrams Press.

D'Ignazio, Catherine, and Lauren F. Klein. 2020. *Data Feminism*. First MIT Press paperback edition. Cambridge: The MIT Press.

Gold, Matthew K., and Lauren F. Klein, eds. 2023. *Debates in the Digital Humanities*. Minneapolis: University of Minnesota Press. https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=3406850.

Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University Press.

O'Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy.* First edition. New York: Crown.

Risam, Roopika. 2015. "Beyond the Margins: Intersectionality and the Digital Humanities." *Digital Humanities Quarterly* 9 (2). www.digitalhumanities.orgdhq/vol/9/2/000208/000208.html.

Waters, Donald J. 2023. "The Emerging Digital Infrastructure for Research in the Humanities." *International Journal on Digital Libraries* 24: 87–102. https://doi.org/10.1007/s00799-022-00332-3.

Resources

- Tapor (https://tapor.ca/pages/about_tapor)
- Voyant (https://voyant-tools.org/)
- Recogito (https://recogito.pelagios.org/)
- Mallet (https://mimno.github.io/Mallet/)