Supporting Material
Digital Methods Final Project Overview
Ariel Hahn | Asa Wilder

Imperial Eyes is a digital humanities project designed and created for Professor Johanna Drucker's *Digital Methods for Research and Scholarship* course in fall 2018. Throughout this course I was exposed to a range of new digital tools and methods which I utilized to build a single, cohesive narrative project. In creating this project I learned how to extract data using an API, use XML and JSON schema, build a website using a static site generator and a range of other valuable skills.

"Imperial Eyes" was a co-creation between myself and Ariel Hahn. I relied heavily on her prior skills and knowledge to guide me through many of the project's technical hurdles and challenges, and I am grateful for her introducing me to github and Jekyll. The following document is a final summary that details our approach, methods, and possible future plans for this experiment.

Project Statement

Imperial Eyes (https://aireuhl.github.io/i_e) uses digital tools to engage with the ideological, physical and geographic creation of the United States. Through a side-by-side visual analysis of historical maps, and a textual analysis of land acquisition treaties, we ask: how did the acquisition and theft of land by European colonizers influence the way the continent was conceived, envisioned and mapped? What can the visual qualities of maps tell us about the role cartography plays in the larger development of colonization? Additionally, using digital tools to draw out deep, hard-to-decipher patterns, we attempt to answer questions about how government treaties helped shape the way land and space were visualized by early cartographers of the American continent. By creating a navigable timeline of maps and treaties, we hope to offer a deeper understanding of the way space is visualized in relation to political circumstances and cultural assumptions. Using a series of 39 historical maps from the New York Public Library's "Maps of North America" Digital Collection, we offer the beginnings of a larger project aimed at recontextualizing the evolution of mapping in America. Paired with interactive text and network analysis of contemporaneous treaties with indigenous peoples, these visualizations encourage digital learners to locate slippery and easily lost historical patterns in an effort to better understand the complex interplay of power, geography and visual representation. This project represents the beginning stages of a much larger-scale endeavor, which unfortunately is beyond our current capacity. We believe that more meaningful and interesting analytics could emerge through further harnessing the power of GIS data and building more links between these two robust categories of data.

Methods and Rationale

Imperial Eyes employs multiple complimentary digital methods to build a cohesive and meaningful display and analysis of visual and textual data. First, using textual analysis tools Voyant and Recogito, we assessed our initial dataset of 30 full-text treatries through informed distance reading and taggable event/person/place mark-up. The text of each treaty within our dataset was scraped from the web by hand and placed directly into an individual .txt file. After the initial readings of and engagement with our textual data, we were able to identify and build a more diagrammatic dataset of names, treaty titles, party information, dates and locations. This data model allowed for a deeper network analysis of 7 randomly selected treatries within our set,

and pointed to overlapping data points (or edges), such as tribal members or United States government officials that were signatories to multiple treaties. Moreover, by geocoding the location names using LatLong.net for this sample, we were able to visualize the geographic relationship between treaty locations. After building our data model and going through multiple iterations of data clean-up in both Google Sheets and OpenRefine, we initially attempted to visualize our .csv dataset through Gephi and Cytoscape. Ultimately, we elected to finalize our network analysis using the visualization software in Palladio. Though the export function within Palladio is not ideal (only a JSON file), this particular software had the lowest learning curve with regard to ingest and provided highly accessible visualizations that are both clear and comprehensive.

To develop our image collection and corresponding dataset for the maps, we again relied on a wide array of methods. We used the NYPL's API to download a set of .xml files in MODS 3.4 schema. A simple Python script was written to scrape these files and download the linked image files. The image metadata was then combined and cleaned in OpenRefine and reformatted into a .csv with a new data model. This data model provided deeper comparative analysis (such as filtering/organizing maps by date or language) and allowed us to add links to the treaty data. By aligning certain fields (date) between our two datasets (treaties and maps), we were able to combine them into a singular visualization in our timeline using Timeline JS .

Moreover, we also offer the beginning stages of a more advanced layer of computer-aided visual analysis. To measure the color, hue, brightness, shape and saturation of each map we used ImagePlot, a suite of high-powered visual analysis tools powered by ImageJ software. The use of computer vision software allows for rapid, large scale visual analysis of dozens of images at the same time. Using this software, we visualized a portion of our dataset using just the images themselves. Through the visualization of large sets of collected images, patterns and trends can begin to emerge that may reveal otherwise overlooked phenomenon.

Community

This project is primarily designed for the academic study of the history of settler colonialism in the United States. Motivated by a desire to highlight the often ignored historical contingencies of

our contemporary political moment, we hope to engage historians and geographers in a wider discussion of the hidden ideologies behind mapping. In its current state, many of our tools are only marginally useful, but with a growing dataset, we are confident more interesting and novel patterns would emerge. These patterns could be useful to students and researchers interested in the intersection of cartography, visual cultural and political science as well as those working across indigenous and native studies, critical geography and information science.

Imperial Eyes was built and designed by two settlers of European descent. Asa Wilder is a third generation American who grew up in traditional Osage territory. Ariel Hahn is a fifth generation Arizonan and descendant of colonial-era settlers who was raised in traditional Tohono O'odham territory. They both currently reside in traditional Tongva territory and are interested in interrogating their participation, both familial and personal, in settler colonialism.

Design

The digital presentation component of Imperial Eyes is a website comprised of two primary sections: one, a comprehensive timeline and, two, a series of visualizations and associated analysis. Additionally, we've included an introductory homepage, a comprehensive next steps page highlighting the potential future of our project, and an about page with information on us as individuals alongside resource links. On the home page, users are met with our custom-designed title/logo as well as a project summary, brief description of our data, and a territory acknowledgement. From here, users can navigate to the interactive, detailed timeline with embedded images and linked text files. The timeline contains records for each of our 39 maps and our sample of seven treaties, including the date issued, names of creators, physical descriptions and a detailed, large-scale image. In addition to the timeline, the second section of our project features a gallery of visualizations and images of analysis tools used to process our data. Users are able to explore the textual datasets of our research through visualizations provided by Voyant and Palladio, highlighting the lingual, networked nature of our data. Then, users can see images of a prototype for an interactive GIS-linked interface. These images, including an animated .gif, demonstrate the possibilities of using warped map images as .kml files in a GIS interface such as Google Earth. While Google Earth has disabled its API and embedding functions, future possibilities for this kind of project remain. Finally, users can

choose to navigate through a gallery of ImagePlot map visualizations. These visualizations order, sort and arrange the maps according to predefined specifications such as hue, color, brightness or date issued. The Imperial Eyes website is designed for both directed, narrative storytelling through the timeline and more open, non-structured exploration through our gallery of descriptive analysis and mapped/networked visualizations.

Technical Specifications and Workflows

Our set of digital assets include .txt files, .csv files, .jpg images, and .kml files. Our .txt files have been created by scraping text off of the web, both manually and through a Python script. Our map images are high quality .jpgs downloaded from the NYPL Digital Collections using a targeted API query. Our .csv files for both sides of the project (maps and treaties) were compiled, cleaned and edited using OpenRefine and Google Sheets. Using ImagePlot, an open source image processing and visualization tool, we produced a series of additional, original image files, as .jpgs. The warped map images used in Google Earth were produced and exported using NYPL Map Warper project. After rectifying each map by assigning multiple linked geo-codes, the .kmg file was produced and exported to Google Earth. To publish these visualizations through the web, screenshots and one animated .gif were produced.

Metadata:

Since this project was small in scope and somewhat iterative in practice, we did not stick to a single metadata infrastructure. Our data models evolved as we experimented with various visualization software until we created satisfactory visualizations. For our map metadata, we relied on the fields created by NYPL's API, which we originally downloaded as .xml Files in MODS 3.4 schema. Initially, we retained MODS date, name, size and location metadata standards. Image-naming was later altered to allow for interoperability with ImageMeasure and ImagePlot. For our treaty data model, our categories we designed for the specific purpose of visualization. We did add a few LOC Name Authority files to our network analysis data, though they were only available for United States authority figures. In our initial search, there were no authority files for any of the native signatories present within our treaty sample.

Digitization standards and workflow:

- NYPL Map images originally digitized as “High Res Tiffs” (2560 px).
    - Physical description of original maps (size and number of composite parts) documented in full maps dataset and on timeline
- .jpg derivatives (1600 px) downloaded for ImagePlot and ImageMeasure processing.
- These .jpgs were then cropped and edited to remove excess border space and rulers.
- Settings for ImagePlot visualizations are highly adjustable and customizable depending on the amount of images included in final image.
- Since sample of treaty images is small, they were downloaded individually and then uploaded to our Google Drive and made public for Timeline JS. A few additional images were manually uploaded to our CMS/Jekyll for use on our website.
- Digitized files of the *Indian Affairs: Laws and Treaties* as evident in our timeline are high-resolution .jpgs, 730 by 900 px at 600dpi.
- Digitized files of original treaties vary in size and resolution with the highest quality being a downloaded image from the National Archives which is 2204 x 3840 px at a less-than-archival 250dpi.

Analytic tools summary and workflow:

- Extracted text from web into individual .txt files, future instantiations of this project would hopefully automate this step through Python.
- Ran .txt files through Voyant and Recogito to perform distance reading and explore taggable event/person/place mark-up.
- Translated elements of .txt documents to .csv using Google Sheets.
- Created an initial data model and cleaned data in OpenRefine.
- Ran .csv file through Gephi and updated data model without success.
- Ran .csv file through Cytoscape and updated data model. Built a network but was not satisfied with visual result.
- Updated and greatly simplified data model. Ran .csv through Palladio, saved JSON file for potential future use and took screenshots of all network visualizations.
- Used NYPL API interface to extract .XML files of map metadata.
- Composed a simple Python code and scraped .xml files for image links.
- Downloaded and renamed .jpg files to allow for ImageMeasure analysis.
- Ran images through ImageMeasure plugin for ImageJ.
- Added resulting visual measurements to .CSV file and converted back to .txt.
- Used .txt measurement file to produce new visualizations with ImagePlot plugin.
- Used Map Warper to rectify and mark geolocations on historic map images.
- Uploaded .kml files to Google Earth instance and saved in “My Places” folder.
- Attempted to render dynamic visualizations of map layers through recorded “tours.”
- Downloaded Timeline JS data model .csv and added map data as appropriated. Uploaded .csv to Timeline JS. Produced initial interactive timeline and upload as an iframe to CMS.

- Added treaty data and created additional descriptions in .csv. Timeline JS automatically updated as did the interactive timeline within our website.

Interface and infrastructure summary and workflow:

Our front-end interface uses Jekyll liquid templating shaped by Forestry.io's block front-matter. Jekyll is a static-site generator that creates easy-to-program blogs and websites. The code is served through and assets held within Github. We experimented with several possibilities – other Jekyll templates, Omeka, Wordpress – before settling on Forestry's CMS. Unlike other Jekyll templates, Forestry can be programmed through your command line, the Github interface, and their own well-designed CMS. This made sense for our project as we were interested in working with code but do not have the technical capacity to create a site on our own, end-to-end.

To build the site, we forked a Forestry.io repository (uBuild) and linked each of our Github accounts with the Forestry CMS page. We experimented with a few Forestry templates, and even did a trial-run where we created all of our own front-matter, before choosing uBuild. Within Forestry, we updated our config.yml and began adding and experimenting with their pre-programmed block-templates. This required us to create each page independently and copy over recurring elements like our navigation and footer each time.

We encountered a few errors throughout the process but were able to fix most issues within the code directly from Github. In terms of shaping the content of our visualization and timeline components, we did the following:

- Exported or embedded necessary visualizations into our CMS/interface after downloading them directly from our software of choice.
    - ImagePlot renderings exported as .jpgs and loaded into our Github repository.
    - Palladio renderings captured through .png screenshots.
    - Voyant and TimelineJS captured through embeddable iframes, which we added directly to our CMS after experimenting with a few workarounds.
    - Google Earth renderings captured through .png screenshots.
    - Embedded maps and treaty images as .jpgs.

We also attempted to use ImageJ's built-in export/embed functionalities to integrate interactive visualizations in HTML.In the end, the embedding function too buggy for implementation. We

complemented our visuals with alt-text to make it as accessible to audiences with limited or zero vision capacity or those interested in engaging with our descriptions in addition to the direct images. The site features no audio or sound video so it is accessible to individuals with limited or zero hearing capacity. Future iterations will also include a search function to provide further access to those who do not engage with the web through scroll and cursor-based exploration.

Sustainability

- To insure the sustainability of our code, we rely on Forestry and Jekyll remaining stable/supported. We can download our code and possibly use the Jekyll liquid templating through our own server in the future should Forestry or Github cease to operate.
- To scale up the project, the current workflow for incorporating new map images and further treaty text needs to be simplified and more automated through Python scripts or other command-line requests.
- A more automated way to ingest content into the data model could eventually be worked into the standard digitization workflow for maps as well as treaties.
- Institutional partnerships offer a promising future for the sustainability of Imperial Eyes:
    - This project would greatly benefit from a partnership with an institution that houses a large collection of historical maps or treaties.
    - Expanding the project would also introduce new voices into our discussions regarding how we model our data, what constitutes as data violence, and what tools make these materials most engaging and accessible.
- The obsolescence of Google Earth's API has already affected the ability to embed functionality as planned. This may happen to the other programs, like TimelineJS and Voyant, that we have embedded directly. It would benefit the project to find a long-term, archival solution for these visualizations beyond screenshots or video screen capture.
- As size and scope of the project increases, Github will no longer be able to function as a viable asset management system. Thinking about other digital asset management systems would be critical to ensure the projects longevity and growth.
- Our TimelineJS plugin relies on images stored in a Google Drive account (this was not initially our plan, but we could not find a way to interface the timeline plugin with files stored in our github repository).
- Creating workflows for data migration would be key to this project's future success. As it stands, we have the capacity to download our repository to our individual computers in addition to leaving it on Github. Another instance of the updated folder would be useful for archival purposes.
- The digital labor required to maintain this project poses perhaps the greatest obstacle to its continued growth and longevity. Without a stable and reliable funding model, the

continued labor necessary to scale up this project is likely untenable. While this initial build did not require a budget for digital asset storage, website design/maintenance or hosting, we recognize that these activities require substantial resources and pose a challenge moving forward.

Summary

Due to time and resource constraints, we are limited to including only a small portion of the land treaties and maps produced during the time period in question. A larger, more in-depth project would build off of this initial set to include not only maps of the whole US or continent, but also more focused and detailed maps of specific local areas, of which there are many more in NYPL's collection. Moreover, by linking the historical maps to geocoded data on contemporary maps, future instantiations of this project could be used to analyze much more minute, and localized historical changes in land visualization and mapping. Ultimately, the goal would be to link the warped map images to the location and network data contained in the treaty data model. This would allow for a single, centralized interface for all of our data and images.

In its current instantiation Imperial Eyes is a prototype and working model for a much larger and more labor-intensive digital project. One of the major limitations to image analysis software such as ImagePlot is that it requires a huge collection of images to produce interesting and insightful data. With only 39 maps in our image collection, the analytics produced by the program do not produce very interesting or meaningful visualizations. Additionally, the treaty text sample in our network visualizations was equally small. A larger sample of adequately modeled treaty data would better interrogate many of the questions we have when engaging with these materials. With a more comprehensive network analysis, it would be easier to see the gravity of how treaties impact they way the United States as we know it evolved. We could see what United States representatives were engaged in the creation of multiple treaties across decades or with multiple tribes, as well as what tribes and specific tribal members were present when the shape of their territory was reduced, moved or shifted by the will of the United States government. Future explorations of these materials would also be complemented by a visualizations and analysis paired with the geo-coordinates we collected.

Beyond the limitations of our current datasets, we also encountered issues with our CMS/Interface that would need to be remedied before continue with the project. Though our Forestry/Jekyll hybrid offers a ready-to-go, code-heavy html website, it doesn't appear to have an accessible .css file to manipulate things like font or padding between content blocks. It is also difficult to build a website based on someone else's template and contend with the legacy data that will sit and, for lack of a better term, rot within your repository. Despite these few issues, we found our chosen repository to work the best for the current needs of our project.

Appendices

Rights Statement:

All treaty images and treaty text are historic government documents and are within the Public Domain. All of the historical map images used in this project were obtained through NYPL's digital collections and were labeled as Public Domain. After initially searching for all maps of North America, the results of the API query were modified to only contain public domain images. Each image is published with the following rights statement on the NYPL website:

> The New York Public Library believes that this item is in the public domain under the laws of the United States, but did not make a determination as to its copyright status under the copyright laws of other countries. This item may not be in the public domain under the laws of other countries. Though not required, if you want to credit us as the

source, please use the following statement, "From The New York Public Library," and provide a link back to the item on our Digital Collections site. Doing so helps us track how our collection is used and helps justify freely releasing even more content in the future.

Data Provenance:

- Map images and metadata downloaded from NYPL digital collections.
- Warped map .kml files produced and exported through NYPL's web-based Map Warper interface.
- All ImagePlot visualizations are original images produced using the ImageMeasure and ImagePlot plugins for ImageJ, an open source visual analysis program.
- Treaty .txt files were scraped from Yale Law School's "Avalon Project: Documents in Law, History and Diplomacy."
- That online text was transcribed from the multiple volumes of Charles Joseph Kappler's *Indian Affairs: Laws and Treaties*, a bound collection of United States governmental laws and treaties with Indian nations that stem from the beginning of this country.
- Images for our treaties are a combination of digitized pages from Kappler's text and actual digital scans of original treaties, gathered from online digital collections and exhibits by the National Archives, University of Wisconsin-Madison, Beaver Area Heritage Foundation, and Oklahoma State University.

Additional Sources:

- Cultural Analytics Lab**:** http://lab.softwarestudies.com/p/cultural-analytics.html
- ImageJ: https://imagej.nih.gov/ij/
- Google Earth Pro: https://www.google.com/earth/versions/
- NYPL Map Warper: http://maps.nypl.org/warper/
- NYPL Digital Collections API: http://api.repo.nypl.org/
- TimelineJS: https://timeline.knightlab.com/