Relocating Rembrandt: Visualizing Provenance Records

Erin Peach
University of Washington
Department of Computer Science and Engineering
epeach@uw.edu

Alice Yang
University of Washington
Department of Statistics
wenhuy@uw.edu

ABSTRACT
Recent focus on looted art restoration has led to governments, museums, and private organizations publishing the provenance for their collections. Little work has been done to standardize and visualize this data could lead to new engagement with the humanities. This project visualizes provenance records for the artist Rembrandt Harmenszoon van Rijn across the dimensions of time and geography. Additionally, this project establishes a standard format for provenance data intended for such visualizations and highlights the challenges manipulating the available data for visualization. Provenance data is manipulated by hand utilizing Excel tools and R scripts assist when applicable. The data is visualized on a world map highlighting the location of individual pieces of art between 1611 and 2016. When a year and piece are specified, the map shows the geographic path the piece has travelled as time progressed. While the time-cost of data preparation is high, the visualization successfully enables the exploration of the movement of art through time at a macro level and highlight times of major geographic change in art. The visualization also enables comparison between the provenances of individual pieces. This project lays the groundwork for data-processing provenance records and creates visualization for art provenance across time and geography at both a large and small scale and lays out recommendations for improved data-processing and art provenance visualization.

Author Keywords
Digital humanities, art provenance, visualization.

INTRODUCTION
During World War II, the Third Reich obtained art pieces from the countries they invaded and shipped these pieces back to Germany, where Adolf Hitler planned to build a Fuhrermuseum in his home town of Linz [7]. After the fall of the Nazi Regime, these pieces were relocated around the world and rarely returned to their original owners. In the last two decades, efforts have been made by governments, museums, and private collections to return confiscated art to its original owners. In 1998, the U.S. Department of State created the Washington Conference Principles of Nazi-Confiscated Art [10], which outlines the best practices of identifying and reclaiming art looted during World War II and calls for the creation of a central registry of suspected looted art, provenance data, and restitution claims. In accordance with these principles, many art museums, like the Metropolitan Museum of Art [8] and the Seattle Art Museum [1], have created digital repositories of art that is confirmed or suspected to be looted. These digital records are publicly available, but are formatted according to the individual institutions preferences. The Getty Provenance Index Database, sponsored by the Getty Research Institute, maintains a central repository of 1.5 million provenance records collected from sales catalogs, dealer stock books, and archival records [9]. While this data is available online and is used to examine the provenance of individual pieces, little work has been completed to manage and standardize this data. Such management is a difficult task as most of the records are kept as transcriptions of handwritten notes, in English, French, German, and Dutch. Additionally, gaps of time and missing information plague the records. This project aims to establish a standardized format for provenance records and to visualize the data at an individual and global level across time and geography.

RELATED WORK
The Getty Provenance Index Database website showcases an example use of the archived information in a visualization of a network diagram showing the auction market in Europe during 1801 - 1820 based on 230,000 records from the database [6]. The viewer is able to interact with this visualization by zooming in and out to better see patterns, trends, and relationships. This visualization and our project look at provenance data at a macro level. However, the data incorporated in our project will focus on the geographic movement of paintings over a large time and for just one artist. This visualization provides a good example for our project as our data also has geographic and network properties.

The Digital Media Lab at the Carnegie Museum of Art is working concurrently on a project, Art Tracks, to establish an ‘unambiguous structure’ for provenance data [2]. As part of a multi-phase project, the researchers at the Carnegie Museum of Art created an installation in their Impressionist Gallery with an interactive visualization which allows museum-goes to choose a piece of art and see its provenance [5].

Our project differs from the Art Tracks project as the researchers at the Carnegie Museum of Art have established

Christ Presented to the People	1400	1865 - 1894	Eastlake, Elizabeth Rigby, Lady. London, England, UK (by inheritance from her husband,
Christ Presented to the People	1400	-1865	Eastlake, Charles Lock, Sir. London, England, UK (d.1865)
Christ Presented to the People	1400	1863 -	Mulvaney (bought at Blamire sale)
Christ Presented to the People	1400	-1863	Blamire, George. London, England, UK (from Smith)
Christ Presented to the People	1400	1844 -	Smith, John. London, England, UK (bought at Harman sale)
Christ Presented to the People	1400	by 1836 - 184	Harman, Jeremiah. London, England, UK (d.1844)
Christ Presented to the People	1400		Emmerson, Thomas. London, England, UK (purchased from Brondgeest)
Christ Presented to the People	1400	1827 -	Brondgeest, Albertus. Amsterdam, Nederland (from Goll van Fra dealer
Christ Presented to the People	1400	-1827	Goll van Franckenstein, Johan. Amsterdam, Nederland
Christ Presented to the People	1400	1734 -	Barij, de (?) (bought at Six sale)

Figure 1. This screenshot shows the "Title", "Accession No.", "Date" and "Owner/Location 1" columns in a .csv file downloaded from the Getty Provenance Index Database before any data processing. The subset illustrates the missing data problem for dates and shows

much stricter guidelines for their provenance records, like accepting only 'a single source of truth for provenance data' and not discarding information [2]. Our project makes no claims to improve or assess the verity of the provenance records. Additionally, information was discarded when it was discovered to be a duplicate. A duplicate record of the painting "Reverend Johannes Elison" was recorded under the title "Mevr Johannes Elison". The duplicated record was removed from the data used in this project. With less rigorous standards of establish verity and accuracy, our goal is not to create a tool that can be leveraged to locate missing pieces or more accurately establish provenance record. Rather, our project hopes to establish a tool to enable the exploration of art across time and geography.

Art Tracks also focuses on the provenance of individual pieces. Our project focuses on the general movement of art over time, by visualizing the global dispersal of a set of paintings between 1611 and 2016. Only after a specific year is chosen, can a viewer highlight a particular piece to see how its location at that point in time relates to its previous locations.

Our project is also location agnostic, as it does not require to viewer to be located near the paintings it references.

METHODS

Data Exploration and Processing

An initial exploration of the downloaded data from the Getty Provenance Index Database showed that the data was mostly transcriptions of handwritten notes. The text was mostly in English but German, French, and Dutch text also appeared. Due to the difficulty of processing such text, the scope of our project was limited to just paintings by Rembrandt Harmenszoon van Rijn, a Dutch painter from the early 1600s. Rembrandt was chosen as the subject of this visualization since he was a prolific artist and many of his paintings are included in the provenance database. The Getty Provenance Index Database contains record for 83 paintings by Rembrandt with a total of 688 data entries. Each data entry represents a specific painting in a specific location.

Due to the difficulty of automatically parsing this data, and the authors' inexperience with natural language processing, most of the data-processing work was done manually in Excel. Built-in parsing tools, like the "Text to Columns" tool, which splits strings into multiple columns using a given delimiter, were used to accelerate this process. R was also used to avoid entering latitude and longitude for all the 546 available locations one by one. The latitudes and longitudes of all 91 cities were obtained and R was used to assign the values to all the 546 entries appropriately.

During cleaning, three common data errors were discovered. These issues are shown in Figure 1.

Ill-formatted Locations

Some owners were listed at multiple locations at the same time. The first location listed was used as the primary location and the other locations discarded. Some locations were recorded in different levels of geographic accuracy (e.g. landmark, cities, counties, states). Locations were trimmed to be city-level for consistency and accuracy.

Missing Locations

Across the 688 total entries, 142 entries did not include location. This situation was handled at a visualization level and required no additional data-processing.

Missing Dates

The dates reported in the data were listed as years. Some entries were missing the beginning date, end date, or both. As the visualization relies heavily on the dates provided in the data, the missing dates were extrapolated. A missing end date would be estimated as equal to the next known location start date. A missing start date would be estimated by the previously known location's end date. In the situation where the end date of entry A and the start date of the next known location, entry B, were missing, the missing dates would be estimated by evenly splitting the time period between the two entries.

Visualization

The visualization was created as a webpage using JavaScript. The D3.js JavaScript Library was used to display a map of the world and place dots to indicate where a painting was located in a specific year. D3.js was also leveraged to include interactive elements, including a form to input a particular year and the list of art pieces which, when hovered over, displays additional information about

Access	Title	Date1	Date2	Code	OWNER	LOC	lat	long	path
34.19	Lucretia	1934	2016			Minneapolis, MN, USA	44.97775	-93.265	1
34.19	Lucretia	1927	1934	2	Jones, Herschel V.	Minneapolis, MN, USA	44.97775	-93.265	2
34.19	Lucretia	1926	1927	2	Reinhardt, and Co	New York, NY, USA	40.71278	-74.0059	3
34.19	Lucretia	1863	1925	3	Carter, J. Purvis	Firenze, Italia	43.76956	11.25581	4
34.19	Lucretia	1853	1862		Burdon, William W.	Newcastle on Tyne, Englar	54.97461	-1.55336	5
34.19	Lucretia	1853	1853		Wombwell, John Cal	London, England, UK	51.50735	-0.12776	6

Figure 2. This screenshot shows the provenance data in the .csv file, after completing the preparation process.

that piece. It was during this stage that missing locations were handled. When information about the location of a piece during a certain time was missing, no location would be displayed on the map for that piece at that time. If a user requested more information about that specific piece, the rest of the available data would be displayed, like owner, title, and the dates during which the piece was kept by that owner.

RESULTS

Data Processing

During the data cleaning, columns of data that were extraneous to the intended goal of visualizing geographical changes in art over time were removed. The following columns were removed: "Artist", "Format/Support", "Comments", "Add'l Subject", "Sale Date", "Sale Note 1", "Sale Note 2", "Sale Note 3", "Sale Note 4", "Sale Note 5", "Notes", and "Copyright". In addition to the data represented in these columns being irrelevant to the visualization goal of this project, all of the columns listed above except for "Artist" and "Format/Support" were extremely sparsely populated – most of completely empty. As the current iteration of this project visualizes data from one artist, the "Artist" column of information was removed. However, were this visualization used with a larger dataset encompassing additional artists, that column should remain in the data.

Following the removal of the extraneous data, the "Owner/Location" columns were parsed as described in Approach, missing dates were extrapolated, and latitude and longitude were added. The cleaned dataset was saved in a .csv format and contained columns indicating the accession number, the title of the piece, the start and end date a single location, a code indicating whether the start date, end date, or both dates were extrapolated, the owner at that location, the text representing the location, the latitude and longitude of the location, and the placement in the order of locations for that piece. Figure 2 shows an example of prepared data.

Visualization

The visualization reads the prepared data and maps the location of art on a map of the world. The version presented in the poster session displayed all the locations art where art

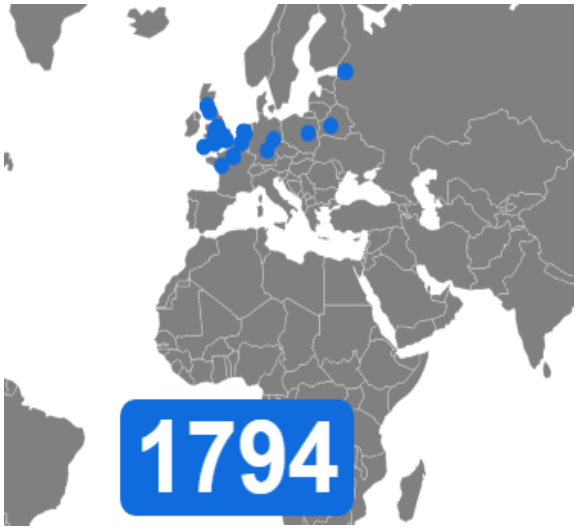

Figure 3. A cropped image of the visualization shows the blue "active dots", which indicate the location of paintings in 1794, the selected year.

was located at any point in time using black circles, referred to as "inactive dots", and the location of art at a particular time in blue dots, referred to as "active dots". Viewers provided strong feedback that this visualization was confusing as the data was highly cluttered. The final version of this visualization shows just the "active dots" to remove extraneous information and focus on the important data. The "active dots" displayed on the map change according to the year queried by the user, as shown in Figure 3.

The viewer is able to interact with the time dimension of the visualization in two ways: a range input and text input. The range input allows the user to see transitions through time as they drag the slider from 1611 and 2016. This interaction is appropriate for exploration and seeing the geographical change at a macro level. Alternatively, a viewer can identify a particular year in which they are interested with the text input form, shown in Figure 4.

Once a year is chosen, a list of the art pieces is displayed to the right of the map. The viewer can hover their mouse over one of these pieces. The "active dot" for that painting in that year remains, while the other "active dots" are removed

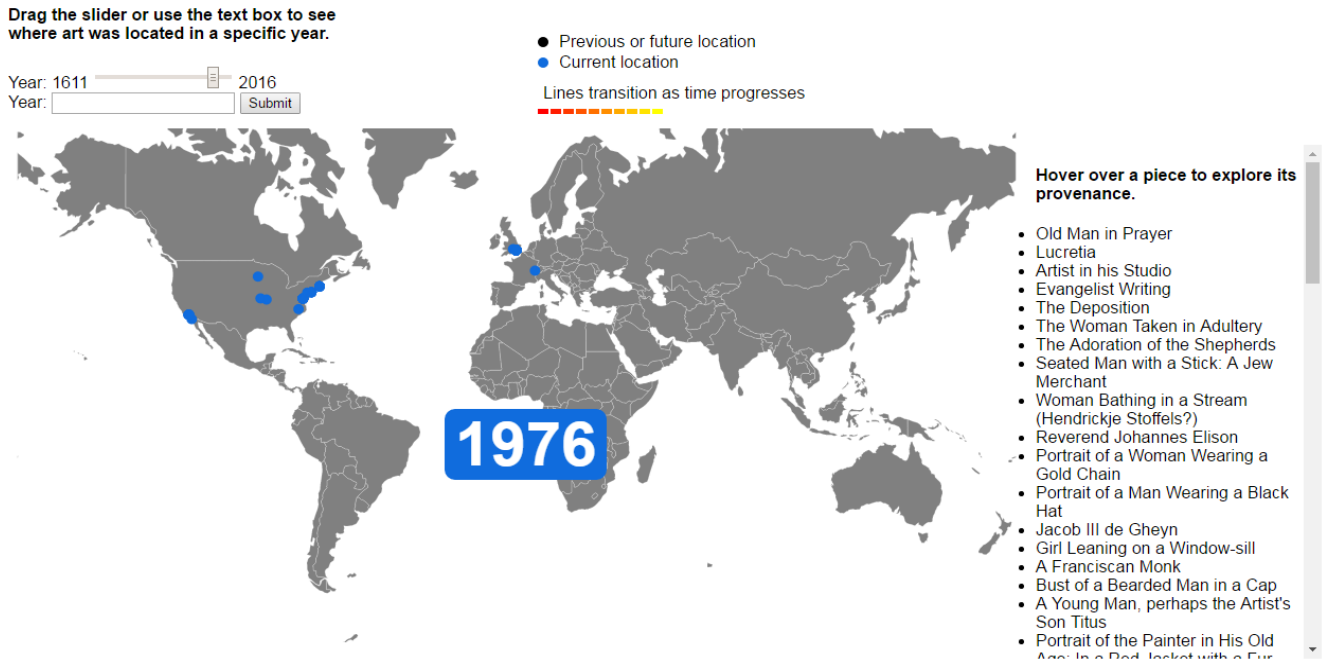

Figure 4. A full image of the visualization shows the input forms, the legend, and the list of pieces with which the user can interact.

from the visualization. "Inactive dots" for the locations that the chosen piece has been in any other year are shown. Both "active dots" and "inactive dots" are connected with lines to indicate the order the paintings were located. The color of each line section is evenly distributed between red and yellow to indicate which transitions between locations appeared more recently than others, illustrated in Figure 5.

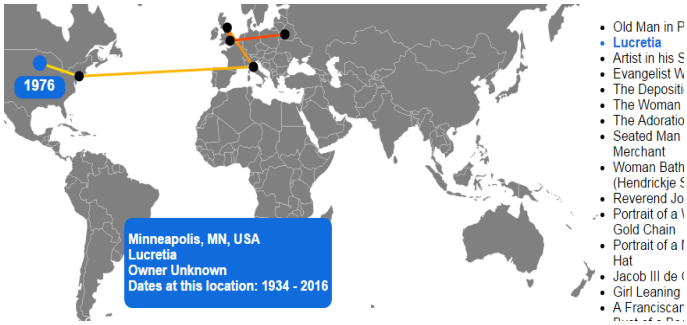

Figure 5. A cropped image shows the provenance path highlighted for the painting "Lucretia" in 1976. The blue "active dot" shows where the painting was in 1976 and the black "inactive dots" show where the painting had been in the past. The red-to-yellow lines indicate the path of the painting as time progressed.

Since the production of machine readable data was the initial part of this project and the data is hosted with the code, the time required to parse the data and produce the visualization is minimal - a matter of seconds. However, since the visualization relies on the browser executing JavaScript commands, minor lag was experienced on slow internet connections.

This visualization uses the data for Rembrandt paintings as an example, but is not constrained to that data. Following the protocols established in this project and using the provided Rembrandt data as an example, a user could procur and clean their own data to be used in the visualization. Due to the complexity of the data cleaning and the need to do most of it by hand, this process would take a considerable amount of time.

DISCUSSION

This visualization is successful in showing the geographic distribution of art over time. As time moves forward from 1611, the "active dots" can be seen clustering in the Netherlands, the home of Rembrandt, and gradually moving to neighboring countries, like France and the United Kingdom. By 1670, Rembrandt pieces have spread as far as Belarus and the United States of America. In 1777 a large cluster of pieces forms in the United Kingdom. During the "Roaring Twenties" in the United States, many pieces of art cross the Atlantic from Europe into North America. By 1970, all but a few pieces have settled in the United States. The visualization allows for such information to be gathered in just a few seconds as the user slides the range input. This feature received the best response during the presentation of this visualization and viewers enjoyed dragging the slider through global changes, like World Wars, to see the effect on art dispersal.

In addition to providing a macro view of art movement, the feature of highlighting a particular piece at a particular time and mapping the path it takes throughout time showed some interesting patterns. Between 1611 and 2016, the piece "Lucretia" travels around Europe before crossing the Atlantic and changing locations in America. However, other pieces remain in the same place their entire existence, like "Seated Man With a Stick: A Jew Merchant" which is only stayed in and around London. In an effort to

understand what might have caused this difference, the authors conducted some quick, casual research which showed the Rembrandt signature on "Seated Man With a Stick: A Jew Merchant" is widely considered a forgery and the painting is now attributed to an anonymous follower of Rembrandt [3]. Enabling the viewer to narrow their focus to just one piece of art at a time allowed for such comparison which resulting in discovering an intriguing narrative about a particular piece.

As the only project addressing a humanities problem, this project drew a lot of interest and got a variety of feedback during the poster session. Many people provided feedback about the initial visualization, where blue and black dots were shown at the same time with multiple black lines overlapping each other. This visualization was initially intended to provide a contrasting view of where art has been, or will be, located and where it is during the requested year. However, people found this visualization too busy and difficult to understand. Many viewers recommended changes, like using small multiples and aggregating lines in the same direction. In the end, the visualization was revised to eliminate the black dots and paths, showing a cleaner and more straightforward look.
Also, people asked about the meaning of the color encoding. This was corrected in the final version where the legend is presented at the top of visualization in easy view.

Finally, small usability changes were requested during the feedback session and incorporated in the final version, like moving the slider in the range input to match the date in the text-input and being able to submit the text-input using the "Enter" key.

There was also valuable feedback not incorporated into the final design due to time constraints. After selecting a year, many users tried to immediately interact with the map, rather than the list of pieces. Some hypothesized that this was due to the list of pieces not looking interactive. Others mentioned that they wanted to explore the art through not map rather than just show the results of outside interactions on the map. Both of these feedbacks could be included in a updated version of this visualization. Formatting the list using a schema that people recognize as interactive would help draw users' attention to the feature. An example of such a schema is the HTML 'select multiple' schema. Additionally, the map can be made interactive by displaying a tooltip when the mouse is moved over an "active dot", which shows the name of the location and the titles of pieces located there at the specified time.

Aesthetically, users wanted to see an image of the piece they specified. With the data on the relatively small scale of just Rembrandt pieces, locating URLs to each of the pieces would not be too time-expensive. However, as this visualization is intended to work for larger data-sets as well, locating as many URLs as necessary for the visualized data is potentially too time consuming.

FUTURE WORK

This iteration of this project laid the foundation for formatting and visualizing provenance records over geography and time. While this iteration focused solely on the works of Rembrandt due to time constraints, an extension of this project would entail processing and visualizing the provenance records of additional artists. In addition to the additional data processing work, it would require the incorporation of attribution to an artist, an additional dimension in the visualization. With more data, the visualization would show more powerful patterns allowing people to draw more convincing conclusions. For a visualization with more data, filters could be applied to highlight the different paths taken by paintings of a particular style, e.g. Baroque, Impressionism.

Before including additional data, it would be ideal to develop a text-processing tool to parse the unstructured handwritten notes and extract the information needed. Without an efficient tool, manually cleaning the larger datasets would be time prohibitive. Developing such a tool might require some knowledge of natural language processing.

Finally, better solutions to deal with the large amount of missing data can be explored. For the unknown dates and location, it could be possible to trace the history of the owner and make a best estimation. More exhaustive research can be done to validate the data.

REFERENCES

1. "About SAM's Collection." Collection Resources. N.p., n.d. Web.
2. Berg-Fulton, Tracey, David Newbury and Travis Snyder. "Art Tracks: Visualizing the stories and lifespan of an artwork." *MW2015: Museums and the Web 2015*. Published January 15, 2015. http://mw2015.museumsandtheweb.com/paper/art-tracks-visualizing-the-stories-and-lifespan-of-an-artwork
3. Follower of Rembrandt. *A Seated Man with a Stick*. Digital image. The National Gallery, n.d. Web.
4. Inscho, Jeffery. "Visualizing the Stories and Lifespan of an Artwork." Web log post. CMOA Blog. Carnegie Museum of Art. Web. <http://blog.cmoa.org/2014/08/visualizing-the-stories-and-lifespan-of-an-artwork/>.
5. Meier, Allison. "Carnegie Museum of Art Makes Its Provenance Accessible and Interactive." *Hyperallergic RSS*. N.p., 04 Sept. 2015. Web.
6. "Network diagram of agents connecting the British, Belgian, Dutch, and French auction markets from 1801–20 using 230,000 records from the Getty Provenance Index® databases". J. Paul Getty Trust and Maximilian Schich. Web.
7. Poole, Robert M. "Monumental Mission." *Smithsonian Magazine*. Smithsonian Institution, Feb. 2008. Web.

8. "Provenance Research Project | MoMA." The Museum of Modern Art. N.p., n.d. Web.
9. Getty Provenance Index® databases. J. Paul Getty Trust
10. USA. U.S. Department of State. Bureau of European and Eurasia Affairs. Washington Conference Principles on Nazi-Confiscated Art. Web.