

A Data Model for Spatial History

The Shaping the West Geodatabase

Evgenia Shnyder¹

The Shaping the West project at the Spatial History Lab created the Western Railroads Geodatabase in order to organize, access, and analyze primary source data. The findings created from the geodatabase can be exported into other mediums to create static and interactive visualizations. The geodatabase allows our project to maintain uniform standards for all of the researchers collaborating on the project and also permits us to limit how researchers access our original data.

Introduction

Shaping the West, directed by Richard White, is a research project at Stanford's Spatial History Lab, which was founded in 2007. It involves mapping nineteenth-century railroads in the American West to study their effect on landscape, the movement of goods, and everyday life.¹ We have created a geodatabase, which we call the Western Railroads Geodatabase, to track movement and its consequences as related to the railroads—whether that movement is track construction, shipment of goods, or worker accidents.² The geodatabase serves as a container that helps us organize, access, and analyze primary source data. It also bridges spatial and nonspatial temporal data to allow for analyses of discrete and seemingly unrelated primary sources, such as historic maps and railroad freight tables. It gives us the power to control how researchers and the general public access our data and to maintain quality control in the project's core.

Data and Organization

The Western Railroads Geodatabase includes both traditional literary sources from the archives, such as letters between railroad company executives, the annual reports of railroad corporations, and labor union newspapers, and spatial sources such as freight tables and historic maps in our research. We can use sources that historians would normally pass over as too dense and opaque or as too hard to merge with more literary data. In the case of our project, the best example of such a source is historic railroad freight tables. Without GIS and computers, no researcher could readily analyze freight tables without spending months manually mapping the station locations and calculating costs among the different stations.

Like other historians, we also use historical maps, but we differ in trying to systematically correlate and merge them with other data. While some of our maps come from atlases or railroad company maps, our most common maps are historic United States

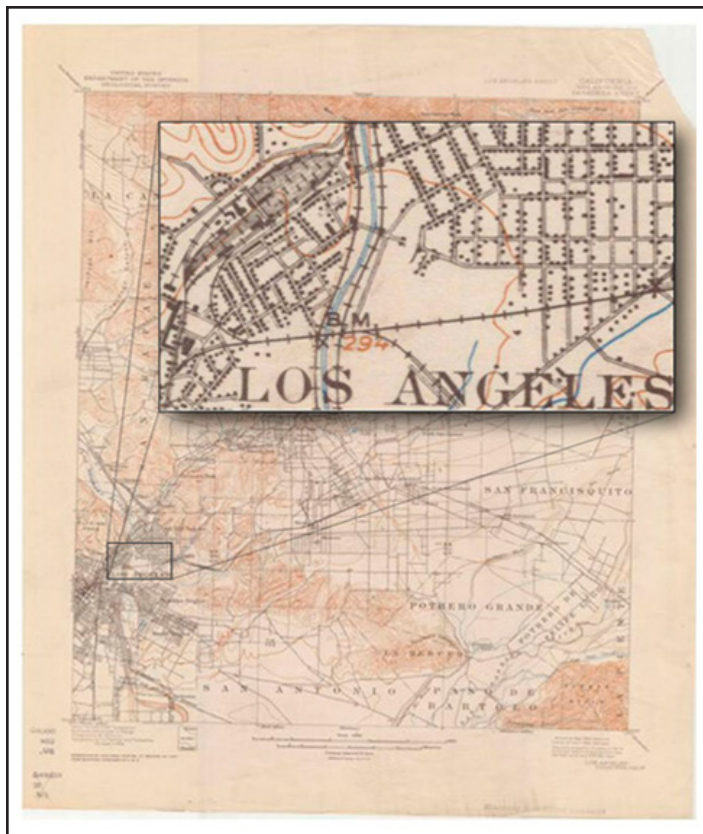
No. 7. CENTRAL PACIFIC RAILROAD.									
SPECIAL FREIGHT TARIFF									
ON GRAIN IN CAR LOADS TO THE GENERAL MARKETS.									
To take effect February 1st, 1876, and remain in force until otherwise ordered.									
For information of employees only and subject to change.									
FROM	TO			FROM	TO			RATE PER TON.	RATE PER TON.
	Sacramento	Stockton	Brooklyn, Oakland, Oakland Wharf, and San Francisco.		Sacramento	Stockton	Brooklyn, Oakland, Oakland Wharf, and San Francisco.		
Redding	6 00	6 00	6 00	Livermore.....	2 80	2 00	2 50		
Clear Creek.....	6 00	6 00	6 00	Pleasanton	3 00	2 20	2 25		
Anderson.....	6 00	6 00	6 00	Sunol.....	3 20	2 40	1 50		
Cottonwood.....	6 00	6 00	6 00	Niles.....	3 20	2 40	1 00		
Buckeye.....	6 00	6 00	6 00	Washington	3 20	2 40	1 00		
Hooker.....	6 00	6 00	6 00	Warm Springs.....	3 40	2 60	1 00		
Red Bluff.....	6 00	6 00	6 00	Milpitas	3 40	2 60	1 25		
Rawsons.....	5 75	6 00	6 00	San Jose.....	3 40	2 80	2 20		
Tylers.....	5 35	6 00	6 00	Decoto.....	3 20	2 60	1 00		
Tehama.....	5 15	6 00	6 00	Haywards.....	3 40	2 60	1 00		
Sesma.....	5 15	5 95	5 95	Lorenzo	3 40	2 60	1 00		
Vina.....	4 75	5 75	5 75	San Leandro.....	3 40	2 80	1 00		

Central Pacific Railroad Special Grain Tariff, 1876. This freight tariff shows the wealth of information contained in all freight tariffs, and the extreme difficulty in analyzing the information without some sort of geographical component.

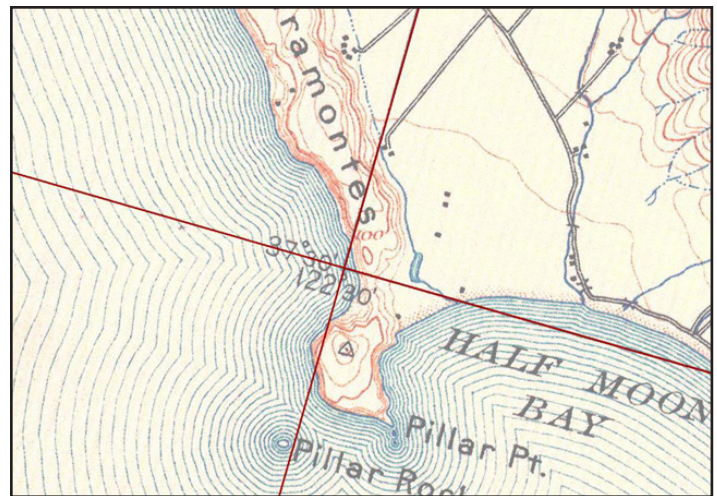
Geological Survey (USGS) topographic quadrangles (quads) surveyed before 1900.³ Historic USGS quads are useful because they were surveyed according to a common standard, have consistent labeling, show railroads in-depth, and cover a large area of the American West, and in particular, California. Each quad, however, covers a relatively small extent of land, and therefore, is not useful on its own for a study of state or national railroad history unless combined with other maps and quads near it.⁴ Researchers could not expect to get a detailed analysis efficiently from historic maps if they had to consistently line them up manually.

We have two forms of digitization in the Western Railroads Geodatabase. The first is merely a form of copying data in which we scan and format primary source materials to create digital versions. For example, after scanning historic maps into the computer, we place the maps in their geographic locations through a process known in ArcGIS as georeferencing.⁵ In this case, we are developing a digitally-formatted replica or medium of the primary source material that can then be manipulated or combined with other sources using a variety of technical tools. We also scan railroad freight tables into the computer and use Optical Character Recognition (OCR) software to transform the document into an Excel spreadsheet that can then be connected to railroad stations in ArcGIS. When we do this, we are repeating exactly what was written on the original document in a digitized form. In both cases, we are mimicking each document in a “feature class” format that can be accessed through the geodatabase.

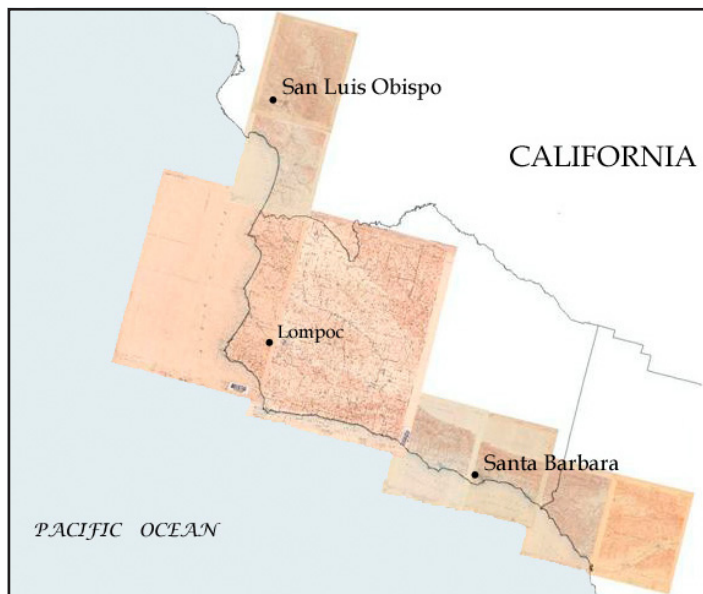
¹Post Baccalaureate Research Assistant, Stanford University Spatial History Lab



Historic USGS Topographic Quad. The Pasadena topographic quadrangle, which maps central Los Angeles, shows the amount of detail within each surveyed quad. Note in particular the amount of railroad detail shown in the inset.

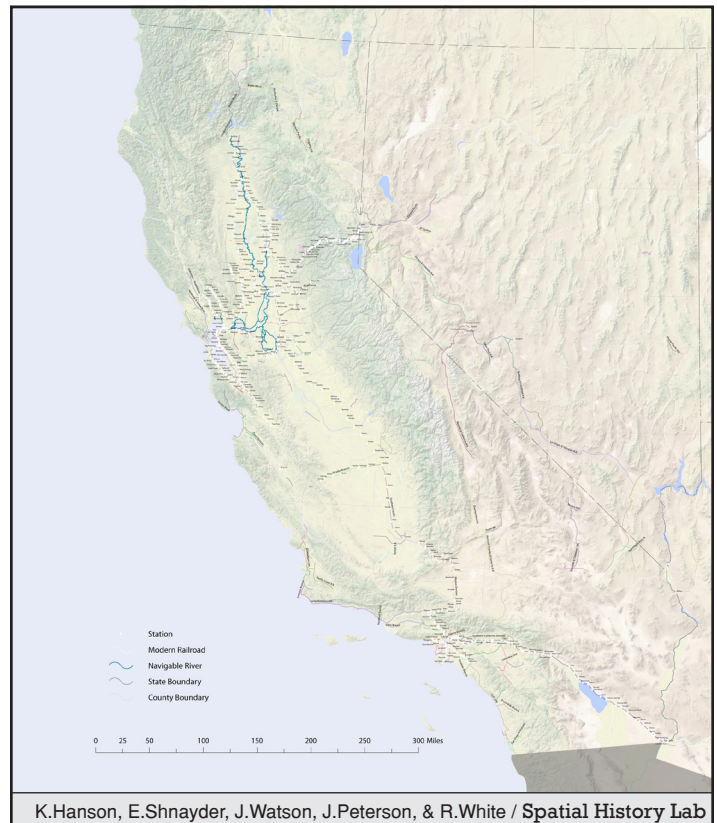


The Georeferencing Process. This close-up of the georeferencing process shows how a historic map that lists its coordinates is placed into “true geographic space” in ArcGIS.



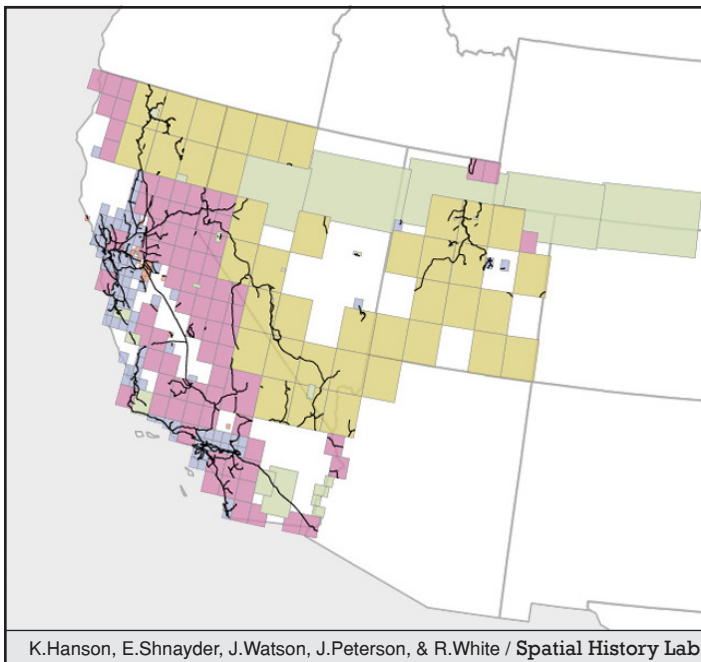
Georeferenced Quads in Southern California. These georeferenced USGS quads demonstrate both the different sizes of quads available to researchers, and the difficulty in trying to find significant amounts of information from one quad alone.

The second type of digitization combines data from two or more digitized primary sources, based on a common “attribute” like a



Full Extent of California Rail Network. This is the extent of the historic rail line and stations traced from historic maps as of the publication of this paper. [Click here for a zoomable version of this map.](#)

time or place, to create new materials. The Western Railroads Geodatabase includes a railroad network created from our digitized railroad lines and stations from a combination of various map and freight table sources. This rail network allows us to track railroad construction and the movement of freight. Once created, our geodatabase can store our primary sources in digitized form so that we can easily access them for analysis and visualizations.



Historic USGS Quad Index. The Quad Index shows the different quads that the Shaping the West project has been able to add to our Western Railroads Geodatabase as of the publication of this paper. Note the range of quads across California—both in terms of breadth and size.

An illustrative example of how the Western Railroads Geodatabase can group sources together and act as an organizational tool is the Quad Index. In an effort to organize the USGS quads once they were georeferenced into ArcGIS, Shaping the West researchers created a new “feature class,” the Quad Index, which shows the quads we have for California and the American West and includes their names, geographic locations, years surveyed and published, and other important information. While we never expected to create a Quad Index, we realized that we required a quick method to search for the maps needed in rail line tracing. The result is as useful as an organizational tool—researchers can quickly locate the quad they need to continue tracing rail line or to analyze data—as it is visually to indicate which quads the Shaping the West project possesses.

Analysis

One of the most important aspects of the geodatabase is its flexibility in adding new data. If the Western Railroads Geodatabase can best be described as a container, then it is also a container that can expand to accommodate new data. Because no researcher can predict what kinds of data will be needed or discovered as the project progresses, a geodatabase must be able to absorb new information and connect it to existing sources seamlessly in order to allow for new analyses. Our geodatabase expands to include new attributes with each set of added data and connects the new attributes with existing attributes. We can thus connect the railroad accidents dataset to the railroad network so that a researcher can map the locations of the accidents. Although the Western Railroads Geodatabase had only been used for storing rail lines and stations initially, we were able to add

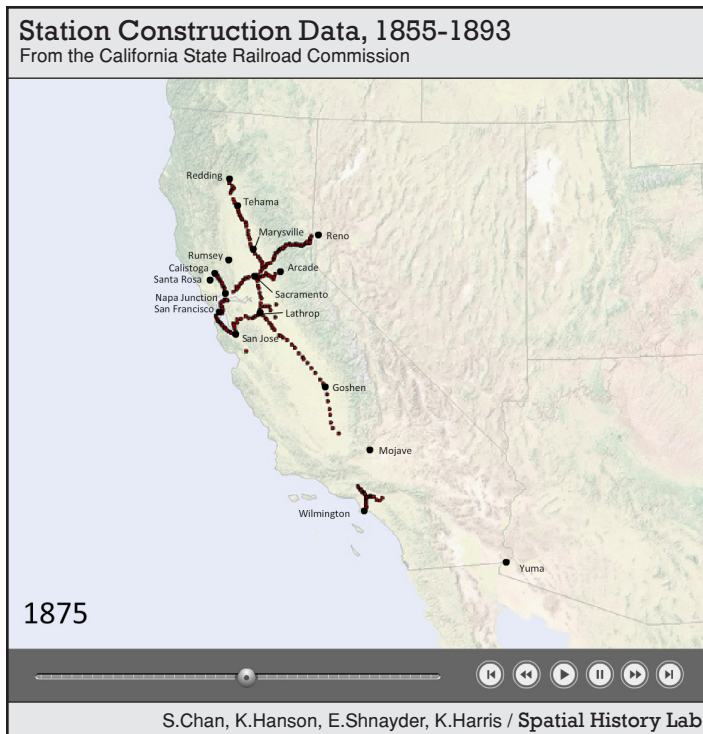
the Quad Index feature class and connect it to our rail network through spatial relationships. We have repeated this process numerous times for other kinds of railroad data, including freight rates, accidents, and station construction dates. We intend for the Shaping the West Geodatabase to remain as a “living” database in the future—researchers will be able to add their own information to our project’s data and the geodatabase will be able to adapt to each user’s research needs.

A geodatabase can also facilitate basic analyses through queries. Once data is entered into the geodatabase, the Western Railroads Geodatabase allows for comprehensive searches through the ArcGIS “find” feature in conjunction with all of the other inputted data sources. For example, we can search the entire geodatabase to see all of the lines operated by a specific railroad company, and then further investigate this company’s history by seeing where they first built rail tracks. A look at the attribute table will tell the researcher, among other things, in what railroad division a section of the railroad line was located and the name of the quad from which the rail line was traced. Such basic queries are often enough to lead to a new line of inquiry.

The Western Railroads Geodatabase allows the Shaping the West project to combine primary sources that would be tedious and time consuming to connect manually. Although rail lines, stations, freight costs, and topography are all innately related in real life, a historian cannot easily piece together the historical remnants of the rail network without the aid of a computer. The geodatabase allows us to visualize rate information spatially.

Our geodatabase also allows us to create visualizations that mimic historical change, such as the growth of the California railroad in the Station Construction visualization. We were able to digitize nineteenth-century railroad station construction dates from the California State Railroad Commissioner into an Excel spreadsheet. We then connected the spreadsheet to the stations in our rail network through ArcGIS and were able to create an animated visualization showing the growth of the railroads in California. By making the connection between space—railroad stations—and time—construction dates—we are able to track temporal and spatial change in the system.

After combining various primary sources together in the geodatabase, we can export them into a different format, such as Adobe Illustrator and Flash, to create new visualizations. Our visualization showing cost-relational space in California best embodies the utility of the Western Railroads Geodatabase. In “What is Spatial History?,” Richard White defines absolute space as the space measured in miles, and relational space as the space measured in terms of cost and time needed for travel.⁶ One of the Shaping the West team’s initial research questions was whether nineteenth-century farmers in the San Joaquin Valley, California were justified in claiming that the railroads charged them higher prices for shipping wheat to San Francisco. Exploring the answer, however, required developing the geodatabase to connect 1876 freight rates to geographic space. The finished visualization, shown on the next page, illustrates absolute space and cost-relational space based on freight rates in 1876 next to each other. By looking



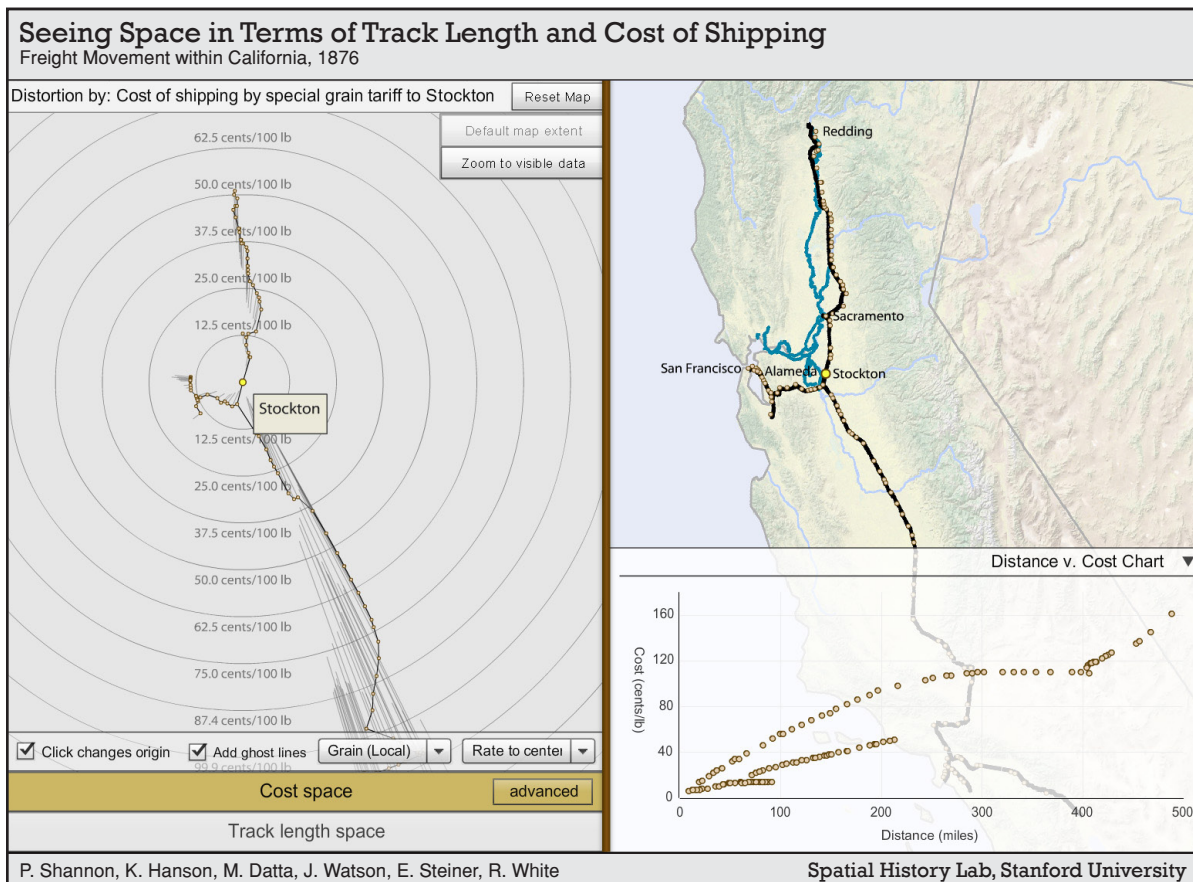
California Railroad Commission, Station Construction Data, 1850-1900. The Western Railroads Geodatabase allows us to track movement. In this visualization, Shaping the West researchers digitized railroad station construction dates and were able to show where and when the California railroad system expanded. [Click here for interactive version.](#)

at the “grain tariff” and “port tariff” in cost-relational space, we were able to show how the railroads manipulated relational space to their advantage and erased the natural advantages that river transport offered. The railroads had maximized their profits and monopolized shipping by forcing farmers to avoid sending their goods to San Francisco through the Stockton ferries.⁷ We could not have done this analysis, however, without building a digital replica of the nineteenth-century rail network and landscape first.⁸

Accessibility

Because the Spatial History Project is collaborative, researchers with various areas of expertise can cooperate to create more nuanced historical analyses.⁹ Although having a large team of researchers work on one project is not always easy, our geodatabase provides a consistent format for our data entry, digitization, and analyses. A single historian could not expect to produce an analysis in as short of an amount of time as a group of researchers working together. Creating the Western Railroads Geodatabase has also allowed us to increase researcher accessibility to our data within the Lab and will eventually allow outside researchers and interested individuals to gain access to our work as well. The geodatabase allows several researchers to simultaneously edit or draw features, which in turn speeds up the digitization or analysis process.

At the same time, however, we recognize that having many researchers work on the same data can lead to a vast array of ideas



Seeing Space in Terms of Track Length and Cost of Shipping.

The distortion visualization shows the power of our geodatabase: it is able to combine geography, the railroad network, and freight tariffs in order to produce a visualization that shows the 1876 railroad system better than any one source could do alone.

[Click here for interactive version.](#)

and approaches at best, and at worst, to chaos and disorganization. The Western Railroads Geodatabase allows us to set accessibility limits for internal or external researchers so that there is a level of accountability and quality control at the project's core. Our geodatabase format ensures that a manager can decide who to grant access to and what that level of access can do. For example, the "railroad editor" can add or delete data, but only the "railroad manager" can specify in which format the data sources are inputted. A "railroad viewer" can access the data to see it, but cannot make changes without the permission of the "railroad manager."

Conclusion

In building a "core" nineteenth-century Western Railroads Geodatabase, we have taken the first step towards producing a new collaborative spatial history analysis, one that is constantly evolving with each added layer of archival information. It is our hope that future researchers become part of the scholarly commons by contributing to our geodatabase. Because we tend to focus on nineteenth-century California, we hope that others will take the time to expand the rail network and analyze railroad history through our "living" database by adding information from their own archive searches. The growth of the railroads transformed American society not only through the economy, but also through perceptions of space and time. Only by studying the movement of the railroads and their influence on life can we begin to better understand the nineteenth-century. Spatial history allows us to do just that by combining a multitude of sources that can then be situated in space as well as in time. For us, no other format has proved as efficient and powerful in doing so than a geodatabase.

End Notes

- ¹ The Shaping the West project defines the "American West" as anything west of the Mississippi River.
- ² Richard White defines "movement" as the key defining difference between spatial history and other forms of history. Richard White, "What is Spatial History?," *The Spatial History Project* February 2010, <http://www.stanford.edu/group/spatialhistory/cgi-bin/site/pub.php?id=29> (accessed September 25, 2010), paragraph 8.
- ³ It must be noted that the first set of professionally-surveyed quads includes maps up to 1915. As such, we have many maps that were surveyed after 1900. As a rule, we always traced rail line from the oldest quad available.
- ⁴ Quads come in a variety of sizes, as small as 7.5' and as large as 60'. Some smaller quads are also layered on top of larger quads from earlier surveys.
- ⁵ Georeferencing in ArcGIS places the maps into their digital geographic locations. This process is essentially the same as a Google Maps team placing street-view photos into their digital geographic locations according to building and street coordinates.

⁶ White, "What is Spatial History?," paragraphs 26, 29.

⁷ This statement and the statement preceding it are better explained in Richard White's forthcoming book, *Railroaded*, from W. W. Norton and Co.

⁸ The freight table values for various stations were connected to the rail network through the attribute value of the station name in ArcGIS. The rail network with all of the included freight data was then exported into Flash.

⁹ White, "What is Spatial History?," paragraph 3.

References

White, Richard. "What is Spatial History?." *The Spatial History Project* February 2010. <http://www.stanford.edu/group/spatialhistory/cgi-bin/site/pub.php?id=29> (accessed September 25, 2010).

Author Information

Correspondence and requests for materials should be addressed to Evgenia Shnyder at eshnyder@stanfordalumni.org.

Acknowledgements

I would like to thank Killeen Hanson and Mithu Datta for their contributions to the earlier versions of this publication. I also owe thanks to Richard White, Kathy Harris, and Jess Peterson for editing various drafts. In addition, Kathy Harris was instrumental in preparing the final images and visualizations for publication.

Rights and Permissions

Copyright ©2010 Stanford University. All rights reserved. This work may be copied for non-profit educational uses if proper credit is given. [Click here](#) for additional permissions information.