

A Data Model for Spatial History: Shaping the West

Evgenia Shnayder¹ & Killeen Hanson² & Mithu Datta³

The history of the nineteenth century American West is the history of the railroads. Where you find the great events, people and places of the first, you are sure to find the straight, iron-girded roads of the second. The railroads were the conduits along which people, traffic, power, money, and influence flowed towards the Western frontier. They represented in the public's imagination, if not always in reality, the great, indelible link to the rest of the world and to the progress of the future.

The Shaping the West project in the Spatial History Lab at Stanford University was established to explore the growth and influence of various railroads in the context of the Western United States. While the eventual goal is to reconstruct a network that includes every historic rail line in the West (i.e. built before 1900), the project places particular emphasis on the railroads of California, especially those around San Francisco. The project aims to connect data gathered from more traditional historic sources (letters, books, newspapers, personal accounts, ledgers) to its specific geographic location using historic maps and surveys georeferenced in ArcGIS. This *spatial history*, as we call it in the lab, examines how historic perceptions of space in the newly settled West were influenced by more than Cartesian geography: specifically by patterns of commerce, communication and cost.

Thus, one early-acknowledged goal of the project was to construct an inter-relational database that would allow researchers to query and explore the geographic links between otherwise unconnected data. Unsurprisingly, at least to those familiar with building a historic GIS, that task proved more difficult than we had initially expected. The challenge of organizing data spatially and temporally (so as to better examine the changes across the West over time) required that we identify precisely what defined a railroad and how and for what purpose we would build the relational links between the physical railroad and any accompanying information. The resulting database allows us to answer the following fundamental questions:

- Can we break down a railroad into discrete spatial objects in GIS layers?
- Can we query by station name or company name and find the correct record?
- Can we locate a station or track length within a larger regional unit, such as county or quad?
- Can we determine the connectivity and adjacency of rail tracks?
- Can we show the changing extent of the railroad over time?
- Can we query associated historic information by location and year and find the correct record?

- Can we link discrete location events and their attributes to existing spatial objects?

The remainder of this paper will explain the Western Railroads Database structure in more detail, explore how this data model allows us to answer the above questions, elaborate on the process of creating the database and acknowledge some of the

1.1 The Western Railroads Database as an Infrastructural GIS choices we made along the way.

The Western Railroads Database was built for two purposes: to function as the foundation for any number of research projects about the nineteenth century American West, and more specifically to explore some of the specific research questions of the Shaping the West project, directed by Prof. Richard White of Stanford University.

Working on the assumption that the railroads are an integral and relevant piece in any history of the nineteenth century American West, it is our hope that our data model will become the foundation for any number of future research projects. Accordingly, we constructed our database to be as flexible and multipurpose as possible. Historical GIS have many advantages, not least of which is the ability to analyze data that has never in new combinations. The Western Railroads Database is, as articulated by Peter Bol in his fine article "CHGIS: Choices Faced, Lessons Learned," an infrastructural GIS.¹ When complete, it will contain the earliest physical locations of all railroads and stations in the West (with the exception of some small, narrow-gauge logging rails), their dates of construction, their locations and participation within historic counties and any changes in name and/or company affiliation.² The database is broad enough to provide analyses at state level and detailed enough to look at railroad growth at the city level.

As an enrichment to the infrastructural framework outlined above, all peripheral information, i.e., information which adds context and content to the rather sparse Western Railroads Database platform, can be joined to the existing feature classes. This information will be research driven according to the interests of individual investigators. County-level census data, for example, once joined to the historical counties dataset, can be used to make rough estimations about the populations, industries and demographics around a specific station or set of stations. Precise information about individual cities can be joined to the stations feature class. Costs and shipment data can be connected to the rail tracks. In this way, the Western Railroads Database will serve as an effective and versatile "vehicle for analyzing a particular body of data."³

We have tried to envision the many different types of data that

¹Research Assistant, Stanford University Spatial History Lab, ²Lead Research Assistant on Shaping the West, Stanford University Spatial History Lab,

³GIS Specialist, Stanford University Spatial History Lab

might one day be joined to the Western Railroads Database and include straightforward ways to make those connections. It must be acknowledged though, that these connections to non-spatial data were shaped most strongly by the more specific research goals of the Shaping the West project, as directed by Prof. Richard White of Stanford University. It is probable that we have not accounted for all possible types of data. In those cases, we attempted to build in flexibility that will allow the user to create the appropriate relationship classes or links that pertain to his or her inquiry.

1.2 Regarding the Structure of the Western Railroads Database

Other large, digital gazetteer projects (such as Peter Bol's Chinese Historical GIS) have successfully used the Alexandria Digital Library Gazetteer Content Standard (ADL GCS) to organize historic information about geographic places. The Western Railroads Database uses a similar organizing strategy as the ADL GCS. Drawing connections between the two data models will help explain our database. The ADL GCS defines the three core elements of a historical Gazetteer as: *place name*, *place location* and *place type*.⁴ In the Western Railroads Database, these three categories translate well to our *place name*, *spatial object* and *external attribute table*. Our relationship classes connect the three. In addition, we have added a date field to historic records when possible.

Developing a computerized model to accommodate real world attributes and features of a transportation network can be challenging because it necessitates combining data with differing spatial attributes. This process entails constructing a system in which points and lines work together as a single unit of analysis rather than two separate entities. The separation of network features into different GIS layers makes it difficult to build that analytical unit.⁵ Accordingly, we built an interconnected network of points, lines and polygons.

The Stations and RailTracks form a transportation network. Conceptually, networks are represented by a set of points and the lines that represent interaction or movement between these points. These can be either physical connections (e.g. rail segments) or invisible connections (e.g., shipping costs between stations). Having a network of stations and tracks allows us to analyze relationships that exist between sets of stations.

1.3 Development of Spatial and Non-Spatial Objects

The basic Western Railroads Database can be broken down into the following types of data:

Spatial: points (stations), lines (rail tracks) and polygons (counties or quads);

Tabular (i.e., Non-Spatial) including, but not limited to: construction data, place names, company names;

Relational: the connections that bridge the gap between the spatial and non-spatial data;

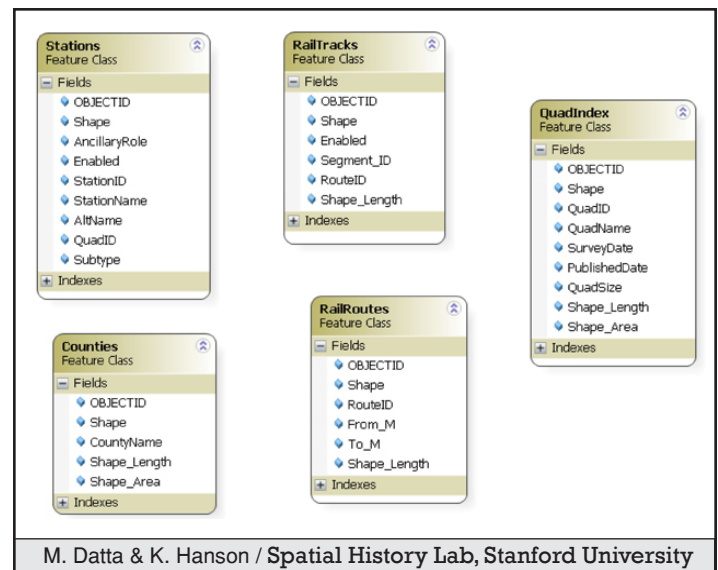


Figure 1 | Feature Classes

1.3.1 Spatial Objects

GIS allows us to mimic real world behavior using objects. From archived historical maps and documents about the railroads we were able to identify the five main spatial objects that make up the infrastructural GIS. Each spatial object is defined by a unique identifier. More detail on their attributes can be found below.

1. Stations (point)
2. RailTracks (line)
3. RailRoutes (line)
4. QuadIndex (polygon)
5. Counties (polygon)

For the purpose of building a working transportation network in ArcGIS, each portion of rail track between stations is an independent line segment. However, in reality the railroads operated in larger “division” units. A set of rail track segments might all be part of the Oregon Division (which stretched from Sacramento to the Oregon border) or the Visalia Division (which ran between Lathrop and Mojave). RailRoutes (a derivative of RailTracks) groups the rail track segments into divisions. The result of this aggregation allows multiple segments (lines) to operate as a single unit of analysis. In GIS this is referred to as Linear Referencing.

An integral component of historical research is source accountability. Because the Railroads Network was digitized from georeferenced Historic USGS Topographic Maps, every digitized object is referenced to an area (polygon) designating the extent of each individual topographic quadrangle. These polygons are collected in the QuadIndex feature class. We used the same technique for other reference regions, e.g. historic county boundaries, census blocks, townships and ranges, etc.

1.3.2 Tabular (Non-Spatial) Data

While the spatial objects above formed the infrastructure of the Western Railroads Database, the *tabular data* in essence hangs on the above infrastructure adding context, depth and substance. The spatial objects and the tabular data work in conjunction to answer particular research-driven questions. In a metaphor used frequently in the Spatial History Lab, one could conceive of the rail tracks and stations as the hardware of the rail system and the tabular data as the software. Thus, we use the hardware of the spatial objects to “run” the software of our tabular data (which might include construction dates, freight rates, census data, etc.) to find answers to our research questions. The types of data can be joined and recombined easily by virtue of their unique ids.

In Richard White’s Shaping the West project, we have used the following pieces of tabular data. This is not an exhaustive list, as we are always identifying new data sources and linking them to our database. This list reflects the needs of our specific project and it is our hope that other researchers will add to this list as they tackle different questions. Each set of tabular data includes the unique identifier for the relevant spatial object. The RouteDetails table, for example, has the RouteIDFK field to make explicit the connection between the RouteDetails and the RailRoute in question.

1. Station Connections
2. RouteDetails
3. RouteReferences
4. Accidents
5. FreightRates
6. Distances

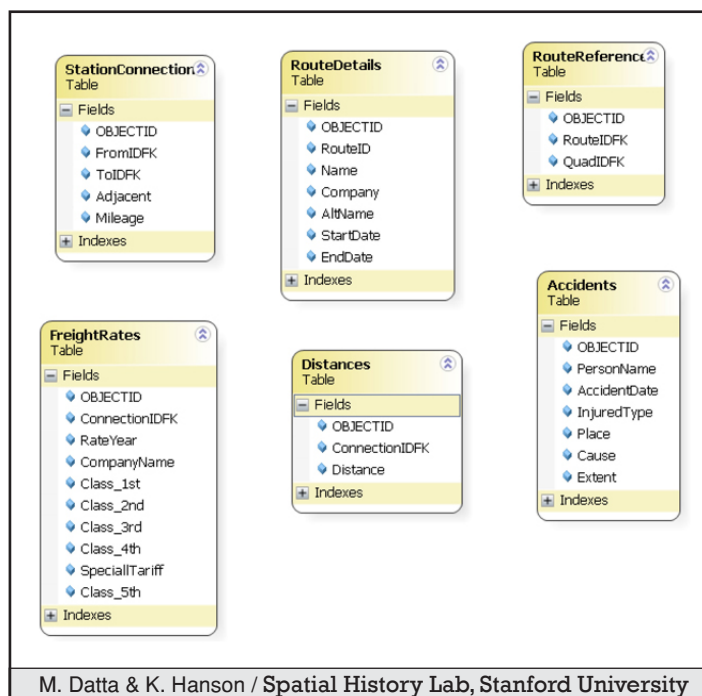


Figure 2 | Tabular Data

Much of the tabular data used in the Shaping the West project deals with the connections and communications between stations. Our tables of freight rates, for example, tell us how much it cost to ship a certain quantity of a given commodity from one station to another. We need to be able to identify connections that exist between a pair of stations and have freight rates attached to this connection rather than to individual stations. For this purpose the StationConnections table was designed. The StationConnections table stores the information of the origin and destination, calculated rail track mileage (the Distances table) and line adjacency. We then tied the freight data as well as similar data that concerning communication or movement between stations to these invisible Station Connections.

RouteDetails, RouteReferences expand upon the RailRoute feature class.

Occasionally, we have information that tells us that a certain event happened at a specific distance down a rail line (e.g., there was an accident along the rail line 30 miles north of the Auburn station). The network we built out of the stations and lines also allows us to find spatial locations for any events that happened not at the stations but at relative distances along the rail tracks.

1.3.3 Relationship Classes

We have the spatial objects and we have the tabular data. *Relationship classes* connect the two and make it possible for us to query the appropriate information and perform analyses on the data. Relationship classes are formalized joins or relates that define the interactions between a set of objects. They are typically defined by a unique identifier found in both the spatial object and the relevant non-spatial tabular data.

The following relationship classes were built to accommodate tabular data specific to the Shaping the West project. They may or may not prove useful to future researchers, but should serve as examples for how to connect spatial and non-spatial data, or two different types of spatial data (as seen in the StationsAreReferencedToQuads relationship).

1. *RailTracksHaveRoutes* - Relationship defined between the RailRoutes object and RouteDetails table.
2. *StationsAreReferencedToQuads* - Relationship defined between Stations object and QuadIndex object.
3. *ConnectionsHaveDistances* - Relationship defined between StationConnections table and Distances table.
4. *ConnectionHasFreightRates* - Relationship defined between StationConnections table and FreightRates table.
5. *ToHave* - Relationship defined between Stations feature class and StationConnections table for the Destination Stations in the StationConnections table.

6. *FromHave* - Relationship defined between Stations feature class and StationConnections table for the Origin Stations in the StationConnections table
7. *RoutesAreReferencedToQuad* - Relationship defined between RailRoutes feature class and QuadIndex feature class

1.4 A Look at the Process of Creating the Railroads Network

This procedure section will detail the steps taken by our researchers as we developed the rail network. Because there are multiple ways to create a dynamic railroad network, the following explanation should be used as a source of ideas and methods and not as a definitive guide.

The rail tracks and stations in the database were digitized from historic maps and are geographically accurate to the best of our knowledge. An essential step, then, prior to digitizing any rail or stations, was to compile and georeference a series of reliable maps that cover the spatial extent of the project: the West. Most spatial objects in the Western Railroads Database are based on historic topographic quadrangles produced by the United States Geological Survey (hereafter USGS quads). The USGS quads were considered reliable for several reasons: 1, they are the work of trained geologists and managed from a single, centralized government bureau; 2, quads are part of a regular, spatially contiguous set in which each edge of a quad abuts another resulting in a largely uninterrupted coverage of the United States; 3, the quads use a consistent and standardized symbology; 4, many of the quads were surveyed at a large scale, with impressive detail; and 5, additional metadata is available for each map, including title, attribution and survey and publication dates. Please note, however, that the earliest USGS quads available in California are from the late 1860s. The USGS has continued to produce and update their topographic quadrangles, but for the specific nineteenth-century focus of our project, we limited ourselves to quads produced before 1915 (both out of preference to the nineteenth-century, and recognizing that after that point many of the rail tracks were moved or removed). When more than one quad exists pre-1915 for the same area, we have preferred the older quad. In this way, we remained faithful to the original locations and names of the railroads and stations.

Occasionally, an historic (pre-1915) USGS quad was not available for a region. In those cases, we sometimes used maps from other sources, or, where we felt confident that the railroad had not moved in the intervening century and a half, we referenced modern rail tracks and maps. Some gaps remain in California and large portions of Nevada. Many USGS quads are already scanned, georeferenced, and available online for download thereby saving spatial historians valuable time in creating a geodatabase.

Depending on the amount of detail desired for a historical railroad project, other map sources are available as well. Before the systematic surveys of the West by the USGS, various military expeditions by the Corps of Engineers U.S. Army mapped the West and recorded the early expansion of the Western railroads. The early surveys serve as good visual references for the rail network on a national scale. Other useful maps for the time period, where

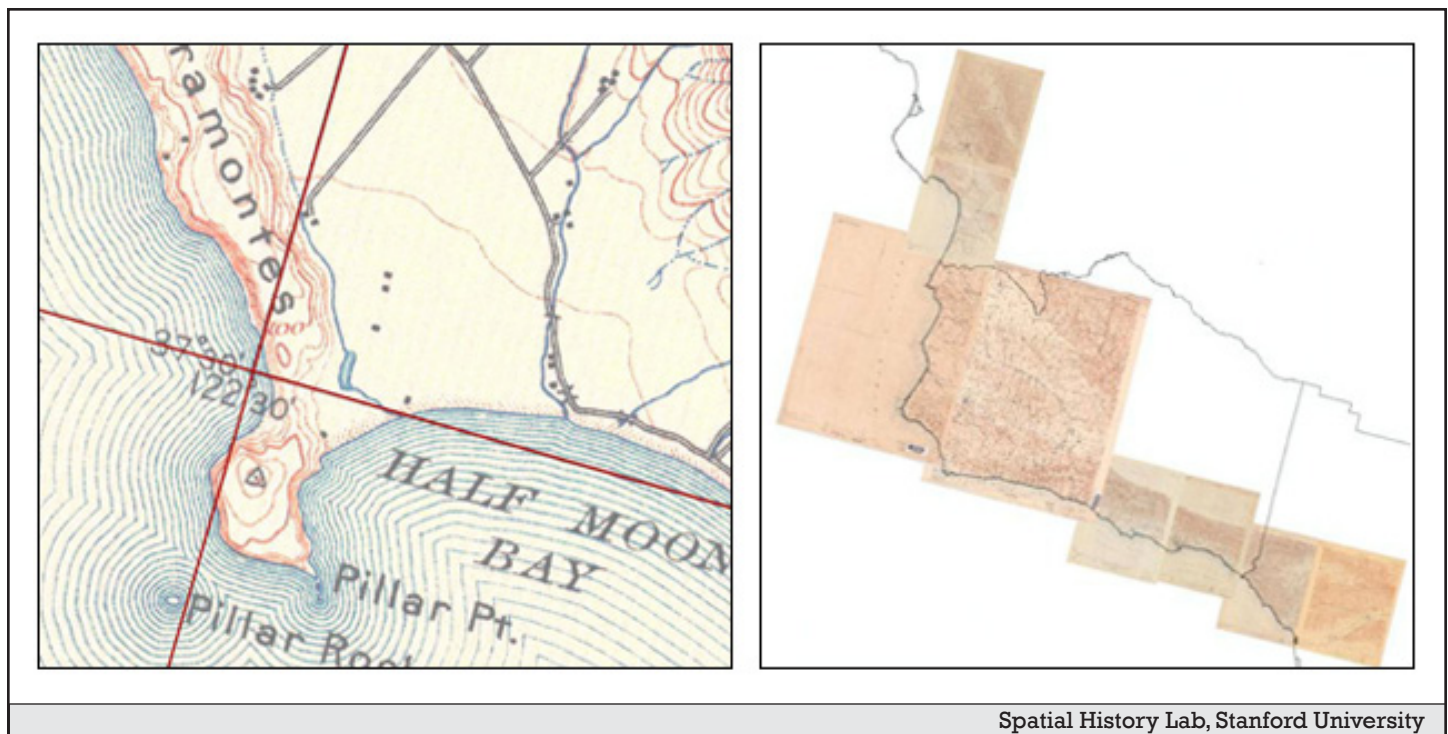


Figure 3 | USGS Quad

quads are unavailable, can be found through Poor's Manuals of Railroads, published annually, Railroad Commissioner's Reports from various states, railroad companies' annual reports, and Atlases of Engineering News and Reports.

Many nineteenth-century maps can now be found online and already georeferenced, such as the Rumsey Collection at Stanford University.⁶ When there are no historic maps available for a certain region, a historical railroad researcher may choose to substitute modern rail (available as shapefiles online from various sources) in place of historical traced rail until a suitable map can be located. In some cases, especially in those which have not experienced a dramatic growth in population since the nineteenth-century, the modern rail is often the same as the historic.

When georeferencing the maps, we found it useful to have a generic reference map of California on screen to speed the process. Additionally, snapping the corners of the USGS quads to a 5-minute grid ensured a precise georeference. When georeferencing USGS quads, we would create an alignment point at each of the quad's four corners and then two more—one each on a horizontal and vertical boundary—and utilize the "second polynomial" option for adjusting.



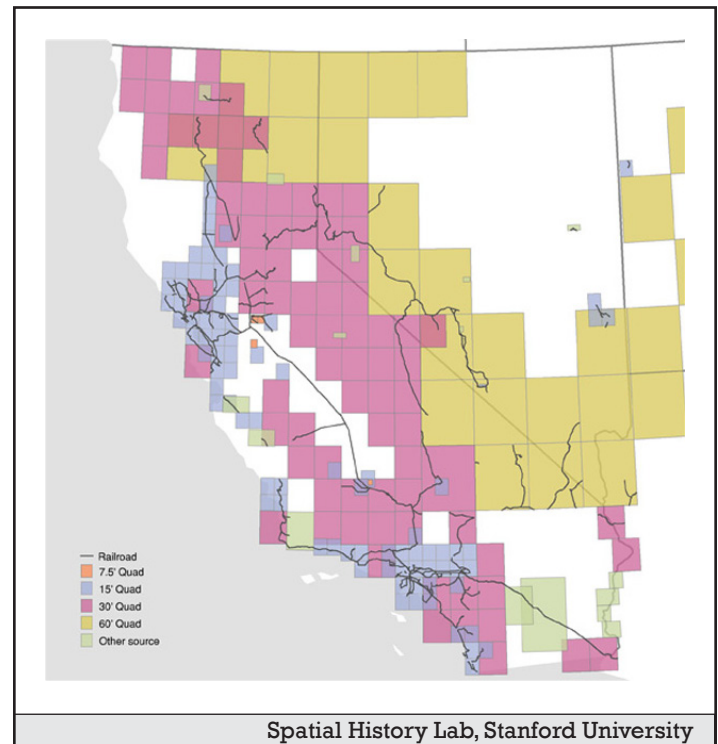
Spatial History Lab, Stanford University

Figure 4 | Georeferenced USGS Quads

When georeferencing the maps, we found it useful to have a generic reference map of California on screen to speed the process. Additionally, snapping the corners of the USGS quads to a 5-minute grid ensured a precise georeference. When georeferencing USGS quads, we would create an alignment point at each of the quad's four corners and then two more—one each on a horizontal and vertical boundary—and utilize the “second polynomial” option for adjusting.

The many USGS quads (the database currently includes 231 separate maps) were organized into the easy-to-use Quad Index, as referenced earlier in section 1.3.1 of this paper. The Quad Index helped us keep track of which USGS quads the project had been able to obtain and georeference, allowed us to search by survey and/or publication date and let us quickly judge which parts of the West we had information for. This is a valuable reference tool for the Shaping the West project and for future historic research projects on the West. The Quad Index has been not only a very valuable organization and time-saving tool but also one of the first attempts to catalogue maps of the nineteenth century West according to their actual location in space.

After georeferencing the necessary USGS quads, researchers would then begin to trace rail line. If the area to be traced was covered by more than one quad, we would first trace from the oldest one.⁷ After finishing all the drawing from the oldest quad, each newer quad (up to 1915) would be loaded. We would trace new rail line not previously found on earlier quads and look for any discrepancies in the line already drawn.⁸ If there were large discrepancies, we checked with our Principal Investigator to decide which rail line trace to keep as the main line. When tracing from 60-minute quads, the Shaping the West project traced at a scale of 1:12,000. For all other quads sizes, the researchers traced



Spatial History Lab, Stanford University

Figure 5 | Historic Quad Index

at a scale of 1:10,000. Researchers were instructed to “break” the traced line, i.e. make complete rail line segments, at each station and rail line intersection. This enabled us to build networks for later analysis.

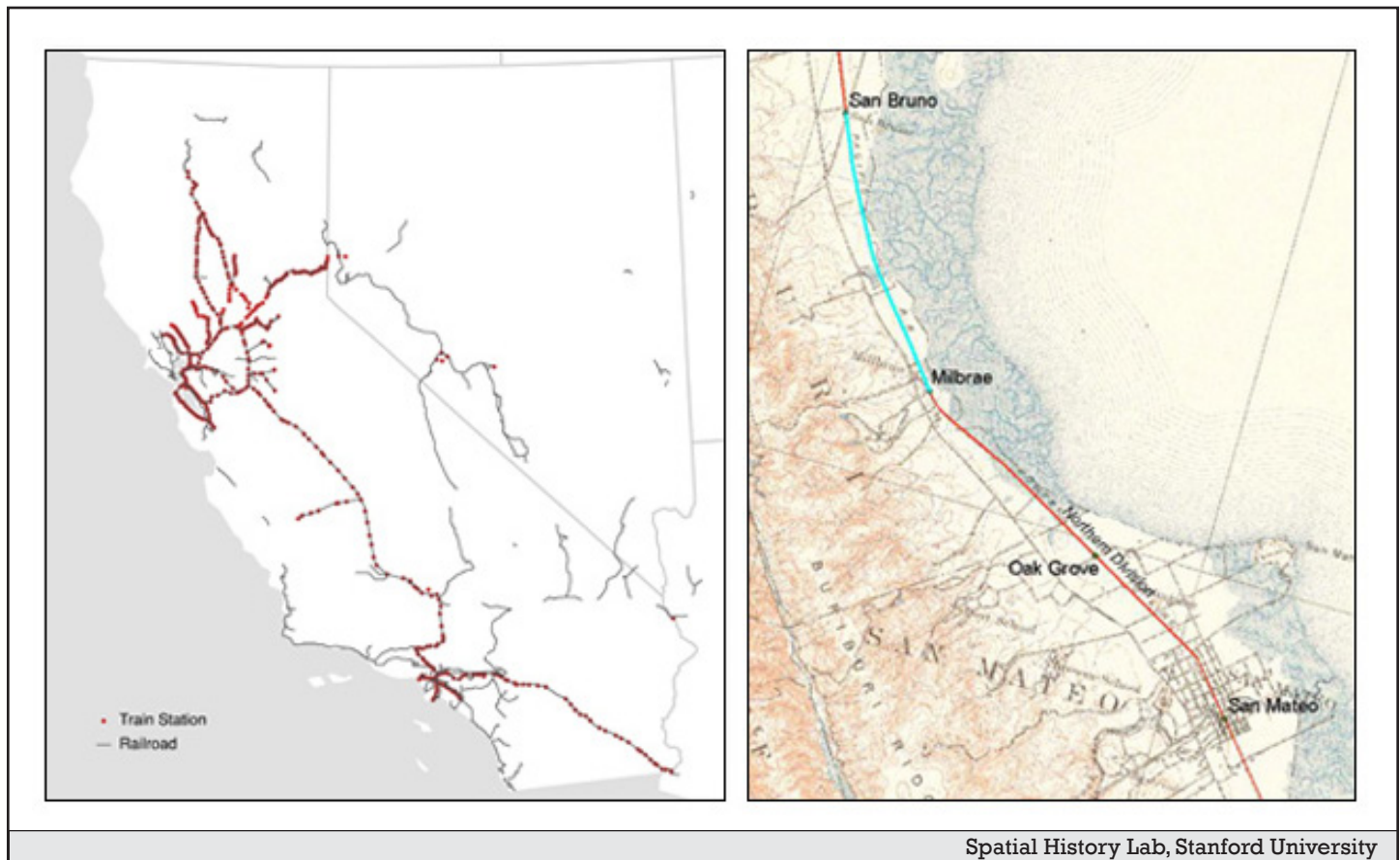


Figure 6 | Traced Railline

After drawing in the rail line, each segment was given a unique identifier and was labeled by name, division (RailRoute) and company. In California, the most common company names were the Central Pacific Railroad, which extended from Northern California to San Francisco and across to Utah, and the Southern Pacific Railroad, which existed from San Francisco down to Yuma, Arizona and across to Texas.

Once the rail line was completed, we added stations. Adding stations is a far more complicated process than just tracing the rail line. We have often found ourselves having to decide between the station's location on the USGS quad and its location (i.e., its distance along the track) on the freight rate tables. Since not all stations are included on USGS maps, especially those which existed as temporary towns when the workers constructed the rail line, our project has also heavily utilized the GNIS domestic database which occasionally gives the coordinate location of a historical town or its historical train station.⁹ Stations were labeled with a unique identifier, a name, an alternate name if the station has ever changed names, the unique ID of the quad on which it was found and, if possible, a construction date.

Through a combination of these three main sources (USGS quads, freight tables and GNIS) we were able to create a rail network of rail tracks and stations.

2.1 Temporal Data

Because the Shaping the West project is a spatial history, we are also concerned with representations of time. The temporal equivalent of this infrastructure/research-driven question is articulated by Peter Bol as a choice between time-series and time-slices.

Ideally, all spatial history projects would be able to show a continuous change-over-time for all of their respective datasets. Peter Bol defines this kind of temporal scope for a historical GIS as a "time-series" and reminds researchers that "a continuous time-series is more true to historical realities."¹⁰ The other temporal option available to spatial historians is a database consisting of "slices-in-time," or certain time periods during which the researcher knows a certain characteristic about the data is true.¹¹ The Western Railroads Database can handle both time-series and time-slice temporal scopes.

Using the Shaping the West project as an example, we see both a time-series (rail track and station expansion throughout California) and time-slices (freight rate distortions). Though not all station construction dates have been found for California, enough stations are labeled to give us a general understanding of the Southern Pacific's and Central Pacific's railroad expansion. A time-series visualization showing the railroads' expansion in California can currently be found at the Shaping the West section of the Spatial History Lab website.¹²

The best examples of Shaping the West's time-slices, and central to one of the project's most important research questions are the tables of railroad freight rates. Freight tables are hard to find and we have complete sets for only a few years. Because the data is so temporally specific, the project cannot examine freight rate change over time. The data is valuable, however, as a snapshot of a single moment or short period of time. We know now, for example, how much it cost to ship grain from Fresno, California to San Francisco in 1876, and can use that information to make conclusions about how the railroads and wheat farmers might have interacted. Time-series and time-slice data are each useful in their own way, only by combining both can the Shaping the West project begin to truly understand the railroads' influence on California in the nineteenth century.

2.2 Challenges and Decisions: Placing the Western Railroads Database Among Other Historical GIS Databases

While creating a coherent and well-organized geodatabase which could be accessible to other railroad researchers was one of the Shaping the West project's early main goals, it was difficult to achieve. The Shaping the West project, like any other historical GIS enterprise, faced difficulties in the process of designing and establishing the current database structure. Some of the problems were unique to our project due to the nature of our research and the sources used—for example, how to determine the “true” spelling of a location which no longer exists and appears differently in various written and map sources. Other problems related to the format and academic direction of the database, however, appear to be more universal to creating historical GIS databases overall. Although Peter Bol of the China Historical GIS (CHGIS) project at Harvard University believes that historical GIS databases will often need to choose between two diametrically opposed options related to the database's purpose and temporal scope, the Shaping the West project has effectively developed a “middle-ground” response.¹³

The Shaping the West project, though largely an infrastructural GIS database, is also research-driven. According to Peter Bol, an infrastructural GIS database is “meant to provide the common basis, the infrastructure, that many different researchers will use in the spatial analysis of their own data.”¹⁴ In fact, our database perfectly answers the three questions Bol asks in relation to determining whether a project's historical database is infrastructural: the project's railroad network's coverage is both broad enough to provide analyses at state levels and detailed enough to look at railroad growth at the city level, the database covers a wide time period showing both stability and change over time, and it is also developed with the types of primary sources to be used in mind.¹⁵ One of the advantages of an infrastructural database is that the researcher never needs to start at ground zero—our database establishes a base upon which new and varying historical sources can be added for more analysis.

As the project has progressed, the addition of more numerical and specific data has allowed the database to develop into much more than just a “base.” For example, freight data records from various years and companies and American Railway Union chapter data which are innately historical, become spatial as well when combined with the project's database. In an ideal situation,

the Western Railroads Database will not need to be reconfigured whenever a new research question or dataset appears. The researcher need only build a connection between his or her new spatial history data and the Western Railroads Database in order to perform analyses. Building a research-driven database on top of an already infrastructural base ultimately creates more leeway and saves time for the researcher.

Sources which initially do not seem to be spatial in nature can quickly become so when combined with georeferenced maps. Freight rates, for example, appear to only be tables with numbers on them, yet paired with special objects we can begin to construct a “cost-relational” space to examine how farmers and manufacturers viewed space. Which stations and cities were closer to each other in this cost-space? Which were farther away? In addition, any data that includes an address can also be included in the geodatabase. A series of monthly lists of secretary addresses for locals within the American Railway Union allows us to map the growth of the Union as well as measure the growing unrest and instability of the organization over time. As of the writing of this paper, the Shaping the West project has been able to combine American Railway Union data, cattle growth, railroad company directors, construction dates, shipping information, railroad accidents and other data sources to the Western Railroads Database for complex and sophisticated analysis.

Our most sincere hope is that this research-driven, project-particular information will become accessible to other researchers, eventually transforming our database into a resource/encyclopedia. In addition to viewing station and rail line locations as established in the infrastructure, researchers will be able to easily and quickly determine through various fields and database relationship links whether the said area also had a railroad union, how much that area paid to ship various goods, what the population of the county was, etc. Access to more information will provoke more questions and enrich research into the nineteenth-century American West.

1. Peter Bol, “The China Historical Geographic Information System (CHGIS): Choices Faced, Lessons Learned” (paper presented at the Conference on Historical Maps and GIS, Nagoya University, August 23-24, 2007), 4.
2. We are using the very impressive and thorough “California Historical Counties Dataset,” available through the Newberry Library and edited by John H. Long, Editor, *Atlas of Historical County Boundaries*. <http://www.newberry.org/ahcbp>.
3. Bol, 4.
4. Guide to the ADL Gazetteer Content Standard, Alexandria Digital Library Project, February 26, 2004. <http://www.alexandria.ucsb.edu/gazetteer/ContentStandard/version3.2/GCS3.2-guide.htm>
5. Harvey J. Miller and Shih-Lung Shaw, “Excerpts from *Geographic Information Systems for Transportation: Principles and Applications*, Oxford University Press (ISBN 0195123948),” http://www.giscale.com/vision/Book/miller_shaw.pdf, 1-2.
6. The David Rumsey Map Collection can be accessed at <http://www.davidrumsey.com>.
7. Tracing was done using the “pencil” tool in the Editor toolbar.
8. A major re-route of the Southern Pacific rail line occurred in the late nineteenth century after a large accident on the Salton Sea.

9. Geographic Names Information System (GNIS), by the United States Geological Survey (USGS), <http://geonames.usgs.gov>.
10. Ibid, 10.
11. Ibid.
12. To see this visualization, visit http://www.stanford.edu/group/spatialhistory/cgi-bin/site/viz.php?id=97&project_id=997. To view other Spatial History Lab visualizations, visit <http://spatialhistory.stanford.edu>.
13. Bol, 2.
14. Ibid, 4.
15. Ibid, 5.

Supplementary Information is linked to the online version of the paper at <http://www.stanford.edu/group/spatialhistory/cgi-bin/site/pub.php?id=23>.

Author Information Reprints and permissions information is available at <http://spatialhistory.stanford.edu/publications>. Correspondence and requests for materials should be addressed to Killeen Hanson (killeenh@stanford.edu).

Rights and Permissions Copyright ©2009 Stanford University. All rights reserved. This work may be copied for non-profit educational uses if proper credit is given. Additional permissions information is available at <http://www.stanford.edu/group/spatialhistory/cgi-bin/site/page.php?id=83>.