

4-2-2008

# A Digital Archive of Historical Railroad Property Valuation Maps

Dean Chauvin

*University of Connecticut*, [dean.chauvin\\_jr@uconn.edu](mailto:dean.chauvin_jr@uconn.edu)

Follow this and additional works at: [https://opencommons.uconn.edu/uccgia\\_papers](https://opencommons.uconn.edu/uccgia_papers)

 Part of the [Archival Science Commons](#)

---

## Recommended Citation

Chauvin, Dean, "A Digital Archive of Historical Railroad Property Valuation Maps" (2008). *UCCGIA PAPERS AND PROCEEDINGS*.  
5.  
[https://opencommons.uconn.edu/uccgia\\_papers/5](https://opencommons.uconn.edu/uccgia_papers/5)

***A Digital Archive of Historical Railroad  
Property Valuation Maps***

2008

*Dean Chauvin  
Department of Geography  
University of Connecticut*

**Abstract:** This paper describes the creation of a digital archive of railroad property valuation maps created for a railway line operated by the New York, New Haven, and Hartford Railroad Company. The original maps were created by the railroad company in 1915 for the Interstate Commerce Commission. The Commission took data on railroad property into account in calculating passenger and freight rates. Property title schedules for parcels along the railway are provided with the maps. Linen copies of close to 3,000 original maps were donated to Archives and Special Collections, Thomas J. Dodd Research Center, University of Connecticut Libraries by the New Haven Railroad Historical and Technical Association. The project to make these archival materials more widely available as digital surrogates involved a collaboration between the curator of the archive and the Dodd Research Center, the University of Connecticut's Map and Geographic Information Center (MAGIC), and Collections Services, Homer Babbidge Library, University of Connecticut.

**Keywords:** railroad property valuation maps, New York, New Haven, and Hartford Railroad Company, historical geographic information systems, rasterization, georeferencing, metadata, Connecticut

**University of Connecticut  
Center for Geographic Information and Analysis  
Storrs, Connecticut**

*Advancing the use of geographic data and spatial analytic techniques  
at the University of Connecticut and in the region it serves*



Copyright © 2008 Dean Chauvin  
All rights reserved.

Cataloging Data

Chauvin, Dean

A digital archive of historical railroad property valuation maps / Dean Chauvin

i, 9, 3, 1 p. : ill. ; 28 cm.

Includes bibliographical references.

(UCCGIA papers and proceedings ; no. 5)

1. New York, New Haven, and Hartford Railroad Company. 2. Railroads—Valuation—Connecticut—Maps. 3. Railroads—Valuation—Rhode Island—Maps. 4. Railroads—Valuation—Massachusetts—Maps. 5. Railroads—Valuation—New York (State)—Maps. 6. Historical geographic information systems. 7. Metadata. I. New York, New Haven, and Hartford Railroad Company. II. Thomas J. Dodd Research Center. III. Homer Babbidge Library (University of Connecticut). Map and Geographic Information Center. IV. Homer Babbidge Library (University of Connecticut). Collections Services. V. New Haven Railroad Historical and Technical Association.

G70.2 .U3 no. 5

Suggested Citation:

Chauvin, Dean. 2008. A digital archive of historical railroad property valuation maps. UCCGIA Papers and Proceedings, No. 5. Storrs, Connecticut: University of Connecticut Center for Geographic Information and Analysis.

Copies of this publication are available at the Map and Geographic Information Center, Homer Babbidge Library, University of Connecticut, Storrs, Connecticut and through *Digital Commons@UConn* at [http://digitalcommons.uconn.edu/uccgia\\_papers/5/](http://digitalcommons.uconn.edu/uccgia_papers/5/)

## **Table of Contents**

Introduction	1
Background	1
Methods	2
Preparation of Templates Describing Individual Maps	2
Scanning the Historical Railroad Property Valuation Maps	2
Georeferencing Scanned Map Images	3
Creating Metadata	6
Searching for Scanned Maps Using CONTENTdm	6
Conclusion	7
References	8
Appendix A – Python Script: Metadata Creator	
Appendix B – Python Script: Bounding Box Coordinate Extractor	

## Introduction

Historical GIS (geographic information systems) is the field of research that employs modern GIS methods to explore past events. The last ten years have seen the burgeoning of work in this field (Gregory and Healey 2007) in disciplines ranging from anthropology (Neubauer 2001) to environmental science (Motzkin et al 1999) and to military science (Juhász 2007).

Some of these contemporary GIS studies concentrate on georeferencing historical data. Georeferencing is the “aligning [of] geographic data to a known coordinate system so it can be viewed, queried, and analyzed with other geographic data” (ESRI 2007). Work in phylogeography (Kidd and Ritchie 2006), museum curation (Wieczorek et al 2004), epidemiology (Guevara 2007), and botany (Kelley 2006) has benefited from georeferencing historical data.

The growth and development of transportation systems including railways has been a focus of work applying GIS to the analysis of historical data (Healey and Stamp, 2000; Gregory and Ell, 2007; O’Kelly, 2007). The availability of data on railway systems is a prerequisite for this research. The purpose of this paper is to describe a georeferencing project to create a digital archive of historical railroad property valuation maps. The creation of this archive will make the data more accessible to researchers interested in analyzing the growth and development of railways and related geographical processes and patterns using GIS.

## Background

In the 1980s, the New Haven Railroad Historical and Technical Association donated an estimated 3,000 railroad valuation maps of the rail line operated by the New York, New Haven, and Hartford

Railroad Company to Archives and Special Collections at the Thomas J. Dodd Research Center, University of Connecticut Libraries. These maps are linen copies of the original 1915 maps created for the Interstate Commerce Commission (ICC) by the railroad. The maps were used to account for the property owned by the railroad and to calculate passenger and freight rates (Pfeiffer 1997). Tables that show the schedules of title for parcels along the railway are provided in these maps. Some maps that represent an area of dense rail lines do not have room for the schedules of title; these schedules are printed separately.

The spatial extent of this collection covers the states of Connecticut, Rhode Island, Massachusetts, and a small portion of eastern New York state. The maps in the collection represent almost all of the railroad line with only a few gaps in coverage.

The maps are frequently used by railroad modelers, railroad and town historians, lawyers, land surveyors, and private property owners. They are stored in rolls containing approximately five to ten maps per roll, sorted by valuation numbers. The rolls are kept in a controlled temperature and humidity environment in the Dodd Research Center.

One mission of the Dodd Research Center is making archival materials available as digital surrogates, where appropriate. A digital collection of these railroad valuation maps would add value to the map collection by allowing easier access by researchers and reducing the handling of the original maps. A collaborative effort by the curator of the archive at the Dodd Research Center, staff of the University of Connecticut’s Map and Geographic Information Center (MAGIC), and staff in Collections Services at the University of Connecticut’s

Homer Babbidge Library produced a set of digital products that will increase the usefulness of the collection.

Successful completion of this project involves meeting three goals. The first goal is to describe each map for the catalog record and to create electronic representations of the maps by scanning. The second goal is to georeference the scanned maps, which enhances the ability to perform useful searches by providing a way to discover the individual maps using their inherent spatial properties. The final goal is to incorporate metadata on the scanned maps into a searchable database. Researchers can use keyword searches on town names and other geographic feature names to search for and discover the scanned maps of interest. The following section describes the methods used to meet these goals.

## Methods

### Preparation of Templates Describing Individual Maps

In order to meet the goals of this project, the maps must first be preprocessed by hand. This step is necessary and involves creating two templates. The first template is used to store important descriptive data of the unique elements contained on each individual map. A second template is created that describes the map elements common to all maps, such as abstract, date of creation, and source.

The two templates are created in Microsoft Excel and are edited to conform to various data recording standards and to make any needed corrections. These templates are later used to create metadata, descriptive information about the data contained on the map. Table 1 lists the descriptive attributes recorded for each of the

individual scanned maps and describes the information recorded for the attribute field.

**Table 1. Attributes of Scanned Railroad Maps**

Attribute	Purpose
Title	Title of map
Filename	Filename of scanned TIFF file
North Orientation	Describes the pointing direction of north arrow
Scale	Map scale
Roll Number	Roll number as labeled by archivist
LCSH Headings	Library of Congress headings
Keywords	Non-Library of Congress keywords
Predecessor Lines	Previous owner(s) of rail line
Valuation number	Valuation number of map
Towns	Town(s) represented by map
States	State(s) represented by map

The maps are preprocessed in batches of approximately 100. Limiting the spreadsheet in this manner creates data sets that are manageable for editing and for integrating with the digital image files of the maps themselves. It also limits load times when data and the image descriptions are uploaded into the database.

### Scanning the Historical Railroad Property Valuation Maps

Once the maps are described as mentioned above, they are scanned. This

task is performed at MAGIC using a Colortrac SmartLF Gx42 large format scanner. The maps, measuring 4.5 feet by 2.5 feet (142 cm by 61 cm), are fed through the automatic loading mechanism of the scanner where they are scanned in black and white at a resolution of 300 dots per inch (dpi). Although these map copies were made around 1915, they are in good condition and no special handling is required while scanning.

The scanned images of the maps are processed in ScanWorks v2.4.4.6 where the images are rotated to the proper orientation. In order to create an acceptable level of contrast for the scanned images, the black and white threshold option is fixed on a setting of 27. This setting removes background speckling while leaving the map details intact. The maps are rotated slightly using the deskew tool to insure they are perfectly square.

The final editing task is performed by cropping the maps to remove the excess collar. Each map has a neatline, the border line drawn around the extent of a map, which is used as a cropping guide.

Once cropping is complete, the files are stored on a computer hard drive. They are saved as an uncompressed Tagged Image File Format (TIFF) images. The finished file size of one scanned map is typically 14 to 15 megabytes.

Figure 1 shows the final result of the scanning process for one of the maps. Note that the north arrow does not always point to the top of the scanned map sheet. Most of the maps in this collection are similarly not oriented with north at the top of the scanned map sheet. When the map in Figure 1 is georeferenced, it is rotated upside down in order to match the source data. This map includes ancillary data, such as road intersections and hydrographic features, that aid in proper georeferencing.

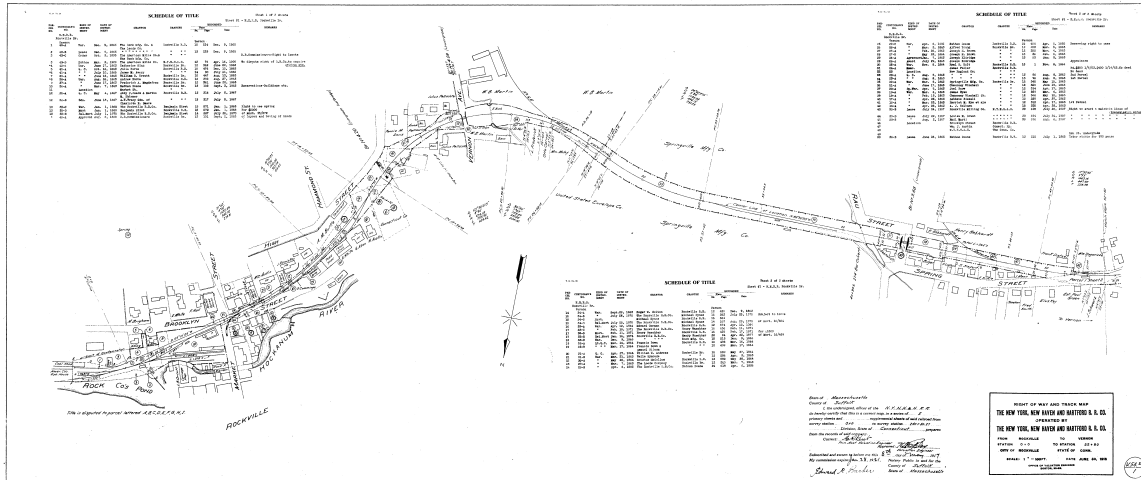
## **Georeferencing Scanned Map Images**

Once the maps have been scanned, they are georeferenced in a GIS. Georeferencing is the electronic manipulation of an image wherein it is rotated and scaled to match other georeferenced data (source data) using a single coordinate system. Georeferenced railroad valuation maps can be searched using their spatial characteristics and interrelationships as well as by using place name strings. Spatial searching is a particularly powerful tool that allows someone to find a map by pointing to an area of interest without having to describe other attributes of the data the map contains.

All georeferencing was performed using ESRI's ArcGIS 9.2 GIS software. Georeferencing requires use of source data; these data have been previously georeferenced. The source data used in this project consist primarily of a mosaic of 1940s-era topographic maps of the entire state of Connecticut. These data are available on MAGIC's Web site (MAGIC, 2008). Although these source maps were created 25 years after the railroad valuation maps, most of the rail lines are present and the maps provide an excellent guide for georeferencing.

Other data are also used to assist in georeferencing. These vector data include road network, hydrography, current rail line locations, and Connecticut town boundaries, which are all available for download from the Connecticut Department of Environmental Protection GIS Web site (Connecticut Department of Environmental Protection, 2007). All data use the North American Datum (NAD) 1983 Connecticut State Plane feet projection.

Georeferencing the images requires a minimum of four control points. Control



**Figure 1.** Example of a scanned railroad valuation map.

points are selected by the user. They are used to match a point on the scanned railroad valuation map to a “real world” point on the source map which already has a spatial reference defined. For example, a road intersection on the railroad valuation map is selected to match the corresponding intersection on the topographic map sheet. Figure 2 shows the corresponding location on the topographic map where the railroad valuation map in Figure 1 should be georeferenced.

Figure 3 shows how the railroad map looks after four control points are selected. The map was rotated to match the underlying topographic map. Note the four sets of red crosshairs indicating the locations of the four control points. Three of these control points were selected where the rail line intersects a road. The fourth point was chosen to coincide with an adjacent railroad valuation map not shown. Once the final set of control points are selected, the user can save the georeferencing information that will allow the image to be displayed properly in a GIS.

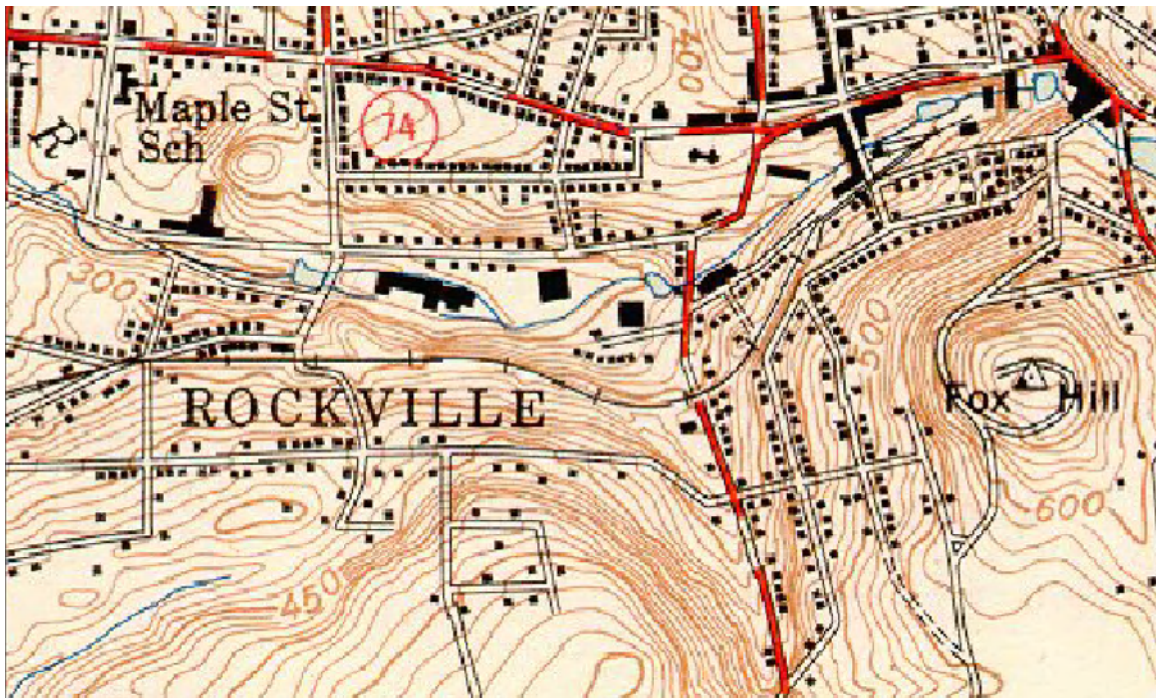
Selecting good control points is required for an image to be properly georeferenced. Intersections of roads, intersections of rail lines and roads, or

distinct water features are all good places to consider control points. Root Mean Square (RMS) error is provided to the user after the fourth control point is selected and represents the degree of correspondence between the different control points. A low RMS error indicates very little transformation was needed to fit the image correctly. The georeferenced railroad valuation maps normally produced relatively low RMS error. Occasionally a map would not produce such a low RMS score. These maps tended to be ones with few, if any, identifiable points that could easily be matched to the source data.

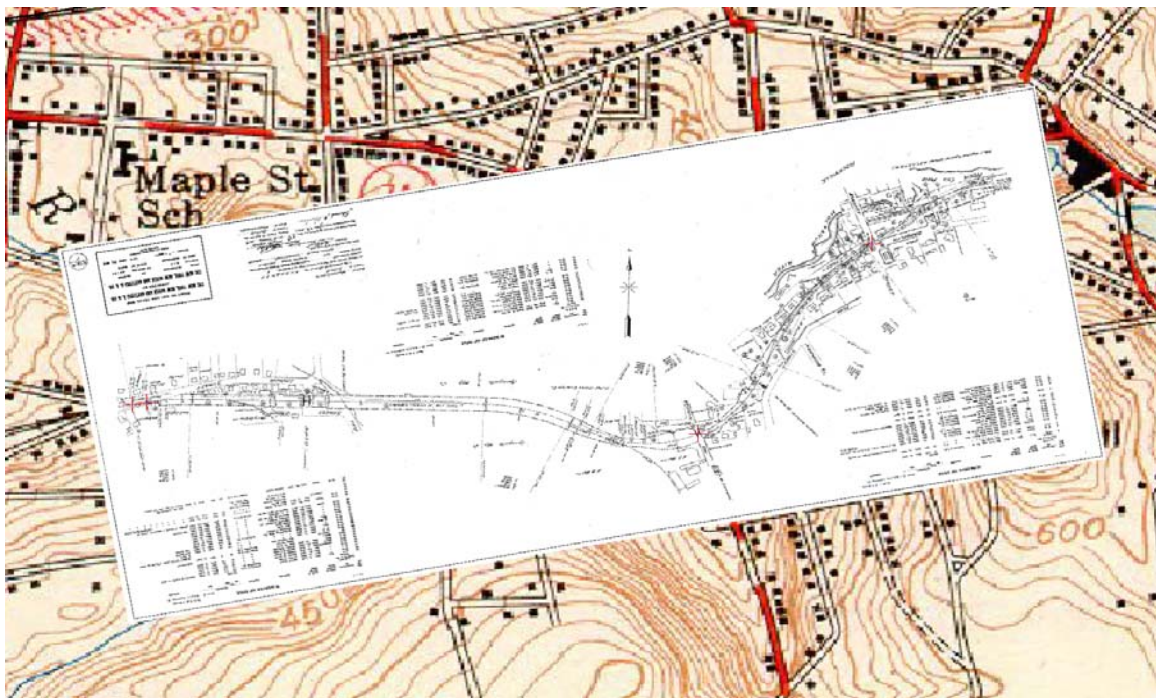
Georeferencing these railroad valuation maps or any other imagery creates additional files that must be managed in order to retain the georeferencing data. The georeferencing process starts with an image file of some sort. In the case of the railroad valuation maps, the image file extension is .tif. This image file is not altered throughout the georeferencing process.

When an image file is first read into ArcMap or ArcCatalog, the software prompts the user to select whether to create pyramids of the image before it loads. This process increases the efficiency of screen redraws of that image. The





**Figure 2.** Topographic map used to georeference scanned railroad valuation map. The topographic map covers the same general area as the railroad valuation map.



**Figure 3.** Overlay of the georeferenced railroad valuation map and the topographic map.

extension of the pyramid file that is created is .rrd. When an image file is first accessed using the ArcGIS software, another file with the extension .aux is created that stores projection, statistics, and other data related to the image.

In addition, two .xml files are written. One .xml file represents the metadata of the image file. When using ArcGIS 9.2 SP2, a file with the extension .aux.xml is also created to store the image transformation data. This file is used in conjunction with the world file (.tfwx) which holds the numerical values used to scale and rotate the image. The .tfwx world file is perhaps the most important file because it defines how the image is rotated and scaled for proper display.

Finally, a .txt file is saved during the georeferencing process that contains the coordinates of the control points. This file allows a patron to rebuild the control points easily if the other georeferencing data are lost or corrupted. In all, six new files are created in addition to the image file for a total of seven files per georeferenced image.

Georeferencing an image does not permanently change the original image file. The original image file will still look the same when viewed in any image-viewing software. The mathematical data used to display the original file correctly are saved in the .tfwx file after georeferencing and are used to transform the image on-the-fly. Rectifying the georeferenced railroad image creates a new raster dataset. This dataset makes the transformation permanent and saves the image as it has been georeferenced. This project does not require rectification of the georeferenced railroad valuation maps.

## **Creating Metadata**

The Federal Geographic Data Committee (FGDC) sets standards for digital spatial metadata in the United

States. FGDC-compliant metadata are created for each georeferenced image. The metadata compiled from the map descriptions, the static metadata, and the .xml metadata file produced by the GIS software recording the spatial information for each map are processed into a complete, FGDC-compliant metadata .xml file. This is accomplished by using a Python script written especially for this purpose. The script supports batch processing many files at once. A copy of the script is included in Appendix A.

## **Searching for Scanned Maps Using CONTENTdm**

Once the railroad valuation maps are scanned properly, the scanned images are transferred to Collections Services at the Homer Babbidge Library. Collections Services loads the scanned railroad valuation maps into the CONTENTdm database, which resides on a library server, together with tab-delimited files created from the spreadsheet templates containing the map metadata. CONTENTdm is a digital collection management software package used to manage digital media, such as images, audio, and video. This software has an integrated Web-based interface that allows for easy searching and viewing of this digital content.

CONTENTdm converts the railroad valuation map images into a different format while importing them into the database. The railroad valuation maps are sent to Collections Services as uncompressed TIFF files and are converted to the JPEG2000 format during the loading process. This conversion is performed to reduce the average size of each image from about 15 megabytes to between approximately three and six megabytes while maintaining a high level of detail. A patron may request the uncompressed TIFF file, which is kept as the archival copy, if needed. The

JPEG2000 image will likely meet the needs of most patrons.

The railroad valuation map images are displayed using a Web-based interface along with the accompanying metadata. Figure 4 shows the interface of the CONTENTdm system. Some of the metadata text is hyperlinked. A click of the mouse on a metadata entry will perform a search of all images matching the criteria in the clicked metadata. Bounding box coordinates will eventually be included in the metadata. These coordinates are extracted from the .xml metadata file created in the georeferencing

process using a Python script written especially for this purpose. Appendix B contains the Python script that extracts the bounding box coordinate data.

## Conclusion

In keeping with the Dodd Research Center's mission to create digital collections of their archival resources, this project has successfully taken a collection of 1915-era railroad valuation maps and converted them into digital form. In addition, additional tasks were performed to increase the usability of the newly

[home](#) : [browse](#) : [advanced search](#) : [preferences](#) : [my favorites](#) : [about](#) : [help](#)

Search UCL Digital Collections

[add to favorites](#) : [reference url](#)

[back to results](#) : [previous](#) : [next](#)

3.6%

**Right of way and track map, the Boston and Providence R.R. Corp. operated by the New York, New Haven and Hartford R.R. Co. ...from Acorn St. to East Providence, station 88+00 to station 123+00, town of East Providence, state of R.I.**

**Title**

Right of way and track map, the Boston and Providence R.R. Corp. operated by the New York, New Haven and Hartford R.R. Co. ...from Acorn St. to East Providence, station 88+00 to station 123+00, town of East Providence, state of R.I.

**Creator**

New York, New Haven, and Hartford Railroad Company

**Type tgm**

Maps

**Subject local**

[Seekonk River \(R.I.\)](#)  
[Rhode Island Company Street Railway](#)

**States**

[Rhode Island](#)

**Towns**

[East Providence \(R.I.\)](#)

**Description abstract**

Between 1914 and 1921 maps and other documents were created by the railroads in the United States in response to the Valuation Act of 1913, which empowered the Interstate Commerce Commission to require an inventory of property owned by the nation's railroad companies to determine the value of their property, the net worth for each railroad, and as a means to set freight rates. The valuation maps available here, dated June 30, 1915, show property owned and operated by the New York, New Haven & Hartford Railroad Company, better known as the New Haven Railroad, which was the predominant railroad line in southern New England between 1872 and 1968.

**Purpose**

The Railroad Valuation Maps show detailed track plans of the New Haven Railroad's right-of-way as well as other property such as railroad stations, roundhouses, yards, and service buildings. The dates the parcels of property were purchased by the railroad or one of its predecessor lines are available on each map. The maps were originally created to account for the railroad company's property, to report to the Interstate Commerce Commission who used this information to establish freight rates. The maps can now be used to account for the history of the ownership of certain parcels of property, to verify property ownership and dates of sale or transfer to the railroad, and as part of an examination into changes in the property previously owned by the railroad.

**Predecessor lines**

[Providence and Worcester Railroad](#)  
[Boston and Providence Railroad](#)

**Roll number**

4

**Date**

June 30, 1915

**Valuation number**

Valuation no. 3, sheet S\_T.4.1; Valuation no. 35.51, sheet 4

**Scale**

1:1200

**Map extent**

1 map : 142 x 61 cm

**North orientation**

North oriented to top right of map

**Format extent digital**

4912.643 KB

**Source**

[Thomas J. Dodd Research Center, University of Connecticut Libraries](#)

**Rights**

Copyright restrictions apply. For more information, contact Archives & Special Collections at the Thomas J. Dodd Research Center, University of Connecticut Libraries.

**Collection**

[New York, New Haven, and Hartford Railroad Company Valuation Maps Collection](#)

**Identifier accession no.**

MSS1998-0378

**Full resolution**

Volume47luconn\_asc\_1998-0378\_map\_35\_51\_4.tif

[add to favorites](#) : [reference url](#)

[back to results](#) : [previous](#) : [next](#)

Figure 4. The CONTENTdm interface.

created digital collection of railroad valuation maps.

The digital images of the railroad valuation maps were coupled with spreadsheet templates containing map metadata and loaded into CONTENTdm. CONTENTdm will allow a patron to query the map metadata and quickly find content that is of interest. The integration of the maps and their corresponding metadata is performed by Collections Services at the Homer Babbidge Library at the University of Connecticut.

The digital railroad valuation maps were georeferenced using a GIS. The processed images can be displayed showing their spatial relationships between one another. The georeferenced railroad valuation images provide additional search functionality. Only a notion of where the map is in relation to towns or other geographic features is needed to find a map.

The georeferenced railroad valuation maps are currently not in a patron-friendly format. A skilled GIS user is required to access the georeferenced maps. In the future, these georeferenced maps should be incorporated into a Web-based interface where a patron could access the georeferenced maps without the help of skilled personnel.

The CONTENTdm records are maintained by Collections Services while the georeferencing data are held by MAGIC, both located at Homer Babbidge Library. If the georeferencing data are published online, it is unclear how the two data repositories will communicate. Another related issue is how the data being served will be stored. If the two systems operate independently of each other, two identical sets of railroad valuation map images are needed. This data storage scheme is redundant and could incur higher costs than necessary for data storage hardware and maintenance. The most efficient manner

to serve this content is to store the data in one location where both CONTENTdm and the georeferencing service can gain access. The design of such a system may be difficult to implement due to current software limitations.

As of this writing, not all of the railroad valuation maps have been described, scanned, or imported into CONTENTdm. Some changes to the template are currently being considered. After the template is final, the maps that have been processed will be available on the CONTENTdm database and will be readily accessible to the public at <http://images.lib.uconn.edu>. Public access should be in place by June, 2008. This access will support the use of GIS with historical data to enhance our understanding of the social and economic landscape.

## References

- Connecticut Dept. of Environmental Protection. 2007. *GIS Data*. Online. <http://www.ct.gov/dep/cwp/view.asp?a=2698&q=322898> [December 18, 2007].
- Environmental Systems Research Institute, Inc.. 2007. Support Center. Online. <http://support.esri.com/index.cfm?fa=knowledgebase.gisDictionary.search&searchTerm=georeferencing> [December 18, 2007].
- Gregory, I.N. and P. S. Ell. 2007. *Historical GIS: Technologies, methodologies and scholarship*. Cambridge University Press: Cambridge.
- Gregory, Ian N. and Richard G. Healey. 2007. Historical GIS: structuring, mapping and analyzing geographies of the past. *Progress in Human Geography* 31(5): 638-653.



- Guevara, Baílde García. 2007. The contribution of ethnography to knowledge on socio-cultural codes related to localized cutaneous leishmaniasis in a health education program in Venezuela. *Cadernos de Saúde Pública* 23(Supplement 1): S75-S83.
- Healey, Richard G. and Trem R. Stamp. 2000. Historical GIS as a foundation for the analysis of regional economic growth. *Social Science History* 24(3): 575-612.
- Juhász, Attila. 2007. A special GIS application – Military historical reconstruction. *Periodica Polytechnica: Civil Engineering* 51(1): 25-31.
- Kelley, Susan L. 2006. Resolving place names in Amdo and Kham: A gazetteer for the Hengduan Mountains region of Southwest China. *Acta Phytotaxonomica Sinica* 44(6): 721-732.
- Kidd, David M. and Michael G. Ritchie. 2006. Phylogeographic information systems: putting the geography into phylogeography. *Journal of Biogeography* 33(11): 1851-1865.
- MAGIC - Mosaic of Connecticut 1940 Era 7.5-min Topographic Maps, 1:31,680. 2007. University of Connecticut Homer Babbidge Library. Online.<http://mapserver.lib.uconn.edu/magic/topo.37800.0000.194-.s31USGS.1.htm>. [December 18, 2007].
- Motzkin G., A. Patterson III, and D. R. Foster. 1999. A historical perspective on pitch pine-scrub oak communities in the Connecticut Valley of Massachusetts. *Ecosystems* 2(3): 255-273.
- Neubaure, Wolfgang. 2001. Images of the invisible-prospection methods for the documentation of threatened archaeological sites. *Naturwissenschaften* 88(1): 13-24.
- O'Kelly, M. E. 2007. The impact of accessibility change on the geography of crop production: A reexamination of the Illinois and Michigan canal using GIS. *Annals of the Association of American Geographers* 97(1): 49-63.
- Pfeiffer, David A. 1997. Riding the rails up Paper Mountain: Research railroad records in the National Archives. *Prologue-Quarterly of the National Archives* 29(1): 52-61.
- Wieczorek, John, Qinghua Guo, and Robert J. Hijmans. 2004. The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. *International Journal of Geographical Information Science* 15(8): 745-767.

## Appendix A

### Python Script: Metadata Creator

```
import os,sys

#####Excel Data Processing#####
#This set of code writes entries to a dictionary using filenames as the key
print "Processing Excel Data. . ."

excelworkspace=r"I:\School\MAGIC\Scripts\metadata\excel_files"#workspace
xmlworkspace=r"I:\School\MAGIC\Scripts\metadata\xml_files"

blankxmlfilelist=os.listdir(xmlworkspace)
filecount=len(blankxmlfilelist)

filelist= os.listdir(excelworkspace)#get all files in workspace

flistgood=[]#empty list containing all relevant files containing transformed XLS files

for f in filelist:#makes list of only TXT files
    if f[-4:]!=".txt" and f[0:3]!="out":
        flistgood.append(f)

exceldata={}
excelscaledata={}

for filename in flistgood:
    infile=file(excelworkspace+"\"+filename,"r")#sets file to variable
    text=infile.readlines()#reads all lines of metadata file

    for line in text:
        pname=line.split("$")
        prefname=pname[1]
        fname=prefname[0:-4]
        prelimdict={ fname: pname}
        exceldata.update(prelimdict)
        scale=str(exceldata[fname][3])
        prelimscaledict={ fname: scale}
        excelscaledata.update(prelimscaledict)
        del fname,pname
    infile.close()

print "Excel Data Processing Complete"

#####Static Metadata Processing#####
#This set of code incorporates all FGDC required metadata stored in staticmetadata.txt

print "Processing Static Metadata. . ."

staticworkspace=r"I:\School\MAGIC\Scripts\metadata\static_metadata"
staticfile=file(staticworkspace+"\"+"static.txt","r")
lines=staticfile.readlines()

staticdata={}

for l in lines:
    hash=l.split("!")
    key=hash[0]
    entry=hash[1]
    minidict= {key: entry}
    staticdata.update(minidict)
    del hash,minidict
staticfile.close()
```

```

print "Static Metadata Processing Complete"

#####XML Metadata Processing#####

print "Processing XML Metadata. . ."

xmlworkspace=r"I:\School\MAGIC\Scripts\metadata\xml_files"
outworkspace=r"I:\School\MAGIC\Scripts\metadata\output"
xmlfilelist=os.listdir(xmlworkspace)#makes list of XML files

counter=0

for fname in xmlfilelist:
    counter=counter+1
    pfnameshort=fname
    fnameshort=pfnameshort[0:-8]
    xmlinfile=file(xmlworkspace+"\\"+fname,"r")
    line1=xmlinfile.readline()#reads three lines into XML file
    line2=xmlinfile.readline()
    line3=xmlinfile.readline()
    mergename=fname.rstrip(".tif.xml")#gets file name used to merge with excel data
    excelinfo=exceldata[fnameshort]
    line3=line3.replace("REQUIRED: The name of an organization or individual that developed the data
        set.",staticdata["originator"])#replace Originator
    line3=line3.replace("REQUIRED: The date when the data set is published or otherwise made available
        for release.",staticdata["datadate"])#replace Pubdate
    line3=line3.replace("REQUIRED: A brief narrative summary of the data
        set.",staticdata["abstract"])#replace Abstract
    line3=line3.replace("REQUIRED: A summary of the intentions with which the data set was
        developed.",staticdata["purpose"])#replace Purpose
    line3=line3.replace("REQUIRED: A summary of the intentions with which the data set was
        developed.",staticdata["purpose"])#replace Purpose
    line3=line3.replace("REQUIRED: The mailing and/or physical address for the organization or
        individual.",staticdata["addtype"])#replace address type
    line3=line3.replace("REQUIRED: The city of the address.",staticdata["city"])#replace city
    line3=line3.replace("REQUIRED: The state or province of the address.",staticdata["state"])#replace
        state
    line3=line3.replace("REQUIRED: The ZIP or other postal code of the
        address.",staticdata["zipcode"])#replace zip code
    line3=line3.replace("REQUIRED: The telephone number by which individuals can speak to the
        organization or individual.",staticdata["contactphone"])#replace contact phone
    line3=line3.replace("REQUIRED: The organization responsible for the metadata
        information.",staticdata["orgname"])#replace organization
    line3=line3.replace("REQUIRED: The person responsible for the metadata
        information.",staticdata["contactname"])#replace contact person
    line3=line3.replace("REQUIRED: The frequency with which changes and additions are made to the data
        set after the initial data set is completed.",staticdata["updatefreq"])#replace update frequency
    line3=line3.replace("REQUIRED: The state of the data set.",staticdata["progress"])#replace progress
    line3=line3.replace("REQUIRED: The basis on which the time period of content information is
        determined.",staticdata["currentness"])#replace currentness
    line3=line3.replace("REQUIRED: The year (and optionally month, or month and day) for which the data
        set corresponds to the ground.",staticdata["calendardate"])#replace calendar date
    line3=line3.replace("REQUIRED: Restrictions and legal prerequisites for accessing the data
        set.",staticdata["accconstraints"])#replace access constraint
    line3=line3.replace("REQUIRED: Restrictions and legal prerequisites for using the data set after access
        is granted.",staticdata["useconstraints"])#replace use constraint
    if excelinfo[5]== "":
        line3=line3.replace("REQUIRED: Reference to a formally registered thesaurus or a similar
            authoritative source of theme keywords.", "None")#replace registered key words
    else:
        line3=line3.replace("REQUIRED: Reference to a formally registered thesaurus or a similar
            authoritative source of theme keywords.",excelinfo[5])#replace registered key words
    if excelinfo[6]== "":
        line3=line3.replace("REQUIRED: Common-use word or phrase used to describe the subject of the
            data set.", "None")#replace key words
    else:

```

```

line3=line3.replace("REQUIRED: Common-use word or phrase used to describe the subject of the
data set.",excelinfo[6])#replace key words

line3=line3.replace("<cntaddr><addrtype>", "<cntaddr><address>"+staticdata["addstreet1"]+"
</address><address>"+staticdata["addstreet2"]+"</address><addrtype>")#add additional
address information
line3=line3.replace(fname+".tif"+"</title>",excelinfo[0]+"</title>")#add title

line3=line3.replace("</mdDateSt></metadata>", "</mdDateSt><eainfo><detailed><enttyp><
enttyp>Scale</enttyp><enttypd>"+excelscaledata[fnameshort]+"</enttypd></enttyp></deta
iled><detailed><enttyp><enttyp>North
Orientation</enttyp><enttypd>"+excelinfo[2]+"</enttypd></enttyp></detailed><detailed><
enttyp><enttyp>Predecessor
Lines</enttyp><enttypd>"+excelinfo[7]+"</enttypd></enttyp></detailed><detailed><enttyp
><enttyp>Valuation
Number</enttyp><enttypd>"+excelinfo[8]+"</enttypd></enttyp></detailed><detailed><entt
yp><enttyp>Town</enttyp><enttypd>"+excelinfo[9]+"</enttypd></enttyp></detailed><det
ailed><enttyp><enttyp>State</enttyp><enttypd>"+excelinfo[10]+"</enttypd></enttyp></d
etailed></eainfo></metadata>")#add additional data

line3=line3.replace("</cntvoice></cntinfo>", "</cntvoice><hours>"+staticdata["orghours"]+"</
hours><cntinst>"+staticdata["website"]+"</cntinst></cntinfo>")#add hours and website
information
line3=line3.replace("&", "and")
line3=line3.replace("\x85", "...")
line3=line3.replace("<br>", " ")
line3=line3.replace("\n", "")
outfile=file(outworkspace+"\\"+fname, "w")#creates empty output XML metadata file

print>>outfile,line1,
print>>outfile,line2,
print>>outfile,line3

outfile.close()

print "Completed writing ",fnameshort," metadata file. File ",counter," of ",filecount,"."

del fname,fnameshort,outfile

print "XML Metadata Processing Complete"
print "Script Complete"

```



## Appendix B

### Python Script: Bounding Box Coordinate Extractor

```
import os

workspace=r"D:\RR_Images\Bounding box extractor"

LFlineList= os.listdir(workspace) #creates empty dictionary of file names

BBList = []

for f in LFlineList:
    if f[-4:] == ".xml":
        BBList.append(f)

output=file(workspace + "\output.txt","w")#creates output file object

for fname in BBList:
    xmlfile=file(workspace+"\"+fname,"r")

    xmlreadline=xmlfile.readline()#reads line 3 of xmlfile
    xmlreadline=xmlfile.readline()
    xmlreadline=xmlfile.readline()

    xmlsplit=xmlreadline.split("<bounding>")#splits data from XML file
    xmlsplit=xmlsplit[1].split("</bounding>")
    del xmlsplit[1]
    xmlsplit=xmlsplit[0].split(">")
    del xmlsplit[0], xmlsplit[1], xmlsplit[2], xmlsplit[3], xmlsplit[4]#deletes unwanted data
    westbb=xmlsplit[0].split("<")#
    eastbb=xmlsplit[1].split("<")#
    northbb=xmlsplit[2].split("<")#
    southbb=xmlsplit[3].split("<")#
    del westbb[1], eastbb[1], northbb[1], southbb[1]#deletes unwanted data

    westbb=float(westbb[0])#pulls values out of a list and converts to floating point
    eastbb=float(eastbb[0])
    northbb=float(northbb[0])
    southbb=float(southbb[0])

    fname=fname.rstrip(".xml")#removes .XML from the end of each file
    print>>output, fname, westbb, eastbb, northbb, southbb#prints data to output.txt file

output.close()#closes file
print "Script Complete"
```