Historical GIS Materials for South Asia Studies in The University of Tokyo

SEKIDO Ippei

Research Fellow, Graduate School of Arts and Sciences, The University of Tokyo; National Institutes for the Humanities
E-mail: ippeman18@yahoo.co.jp

Abstract   The purpose of this paper is to introduce the historical GIS-related materials in TINDAS[1] and to provide access to the basic materials to those who intend to incorporate GIS into their research. In addition, this paper intends to promote the use of GIS in various research fields by introducing a method of converting data into GIS by referring to historical research now underway.

Key words   historical GIS, India, South Asia, historical maps, database

I.   Introduction

Historical GIS (Geographical Information System), which is a research field that integrates GIS into historical research, has been developing since the late 1990s. TINDAS has also adopted historical GIS as an important analytical method to promote research on long-term economic developments and historical changes in South Asia. In recent years, the development of GIS software has made it easier for non-geographers to introduce GIS into their research. However, in the field of history or fieldwork-based research, there are still many difficulties in the introduction of GIS because of limited data. In addition, it takes a lot of work to convert their data into GIS-based datasets. The purpose of this paper is to introduce the historical GIS related materials in TINDAS and to present access to the basic materials to the researchers who intend to incorporate GIS into their research. In addition, a method of assigning geographic information to the dataset that the researcher collects will be described with specific examples.

## II.   Historical GIS-related materials in TINDAS

TINDAS possesses a rich collection of historical materials from British India that were collected mainly by Mizushima, the former director. These catalogs are available on the TINDAS website.[2] Of these materials, there are two GIS databases that provide the research infrastructure for our historical GIS studies.

One is a database of the polygons of the district-level administrative boundaries of each census. It is an indispensable base map when we analyze census data in GIS. India has a proud tradition of conducting censuses con-

tinuously almost every 10 years since 1872. The district-level data are well-developed for the census. However, because the administration boundaries of districts changed frequently as districts were carved, consolidated, and reorganized, it is necessary to reorganize the data in accordance with the transition of the district boundaries when comparing census datasets from different years (Figure 1). This takes a lot of effort. By using the polygons on GIS, the census datasets can be easily compared through time. Its utility as a base map is very high in India, where the census has a long and great history. These shape files were purchased mostly from ML Infomap, a company that develops and supplies digital maps of India. As for other databases of polygon shape files, the GADM (Database of Global Administrative Areas) provides the shape files of administrative boundaries of almost all the countries in the world on their website,[3] and it is freely available for academic and other non-commercial use.

The other is a database of place names and their locations extracted from historical maps. It covers almost all of India using maps from different eras from the late 1800s to 2000s. The place names and their locations are acquired from topographic maps in the scales of 1/63,360 or 1/50,000. In cases where the next maps of 1/63,360 or 1/50,000 are unavailable, maps of smaller scales are used. These are now searchable on the website of India Place Finder,[4] described in section **III** below.

In addition, we would like to introduce the following three kinds of geographical datasets that have not been converted into a GIS dataset or a digital format: historical maps including topographic maps in the scales of 1/63,360 and 1/50,000, the sources of which were the India Place Finder, administrative atlases from 2001 and 2011, and a series of satellite photographs by CORONA spy satellites.
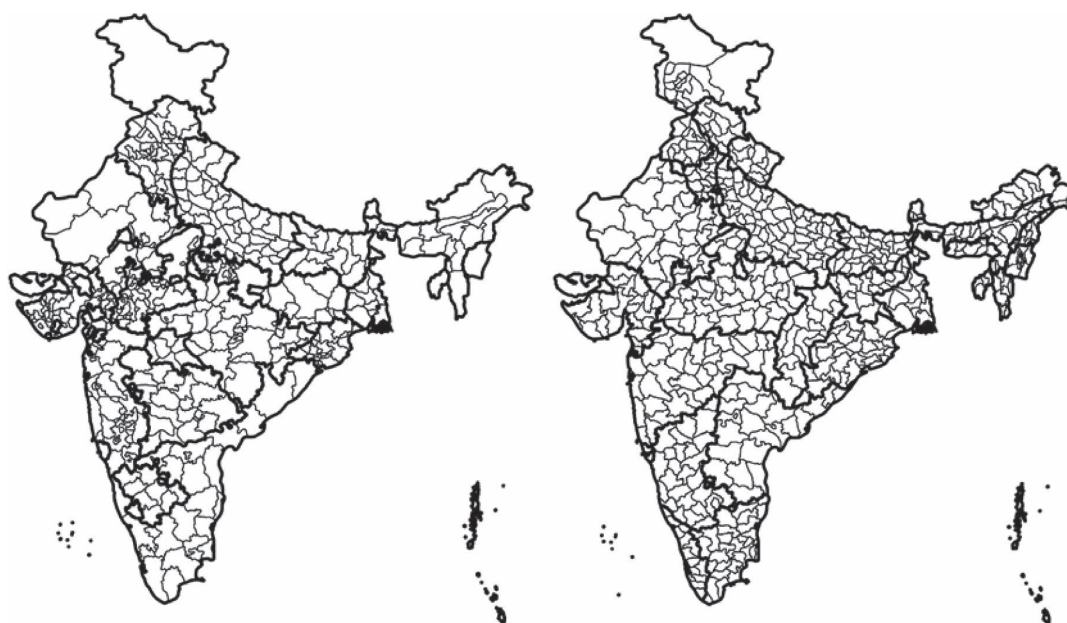
Figure 1. Boundaries of districts: 1881 census (left), 2001 census (right)
Source: ML Infomap.

As for historical maps, TINDAS owns more than 4,000 maps of the survey of India on a scale of 1/50,000 (or 63,360). In addition, 6,000 maps, including city maps, smaller-scale maps, and other maps, have been digitized. As mentioned above, the place names and their locations on the survey map of India have already been added to the GIS database. The historical maps, however, contain much more information: topography (elevations, rivers), railways (bad gauge and other gauge), roads (metered with mile-stone, unmetered, cart-track, camel-track, mule-path, and footpath), post and telegraph offices, police stations, land uses, and so on. It is valuable to be able to observe the distribution and development of these features. In addition, the map helps us imagine the landscape of the site at the time. It is important for not only the historian, but also the researchers who do the fieldwork for current issues, to recognize the landscape of the study site in the first half of the 20th century. Although a catalog of these maps is not currently available on the TINDAS website, a new system is being developed to facilitate the retrieval of these historical maps.

TINDAS owns all administrative atlases of 2001 and 2011 census issued by each state. The main contents of these atlases are sub-district (Taluk and Tahsil) level maps. The maps also show the boundaries of villages, the smallest administrative units.[5] The status of cities and towns, including the census town, outgrowth, municipal committee, and municipal council, is also shown. The boundaries of panchayat unions and revenue circles are also expressed in bold lines. Information other than administrative units, including national/state highways, railroad lines, rivers,

and streams/canals, is available from the maps. Apart from the latest maps, administrative changes in districts and sub-districts from the previous census are shown on the map, which is very easy to understand. In addition, the history of district composition in the past census is summarized. Haryana's administrative atlas of the 2011 census, for example, shows district boundaries for each census every ten years from 1872. These are useful materials to grasp the basic geography and administration of each area.

CORONA satellite photographs were captured by the United States spy satellite project in the 1960s and 1970s. They were released in 1997 and are available for purchase. TINDAS owns photographs from 1964 to 1968. Although in black and white, these photographs have a relatively high resolution and can distinguish land use. TINDAS is now creating its catalog including data on shooting time and shooting range.[6]

III.   India Place Finder

This section takes a closer look at India Place Finder. India Place Finder is a tool to find geographic information for any alphabetical place name in India. Place names change with time, and names written in the original languages can be written in many ways in English. Even the place names written in academic works are not always accurate. This system searches for candidates that resemble the input alphabetic notation and displays them on a map to help users obtain appropriate locations from place names that contain some fluctuations (Figure 2). By com-
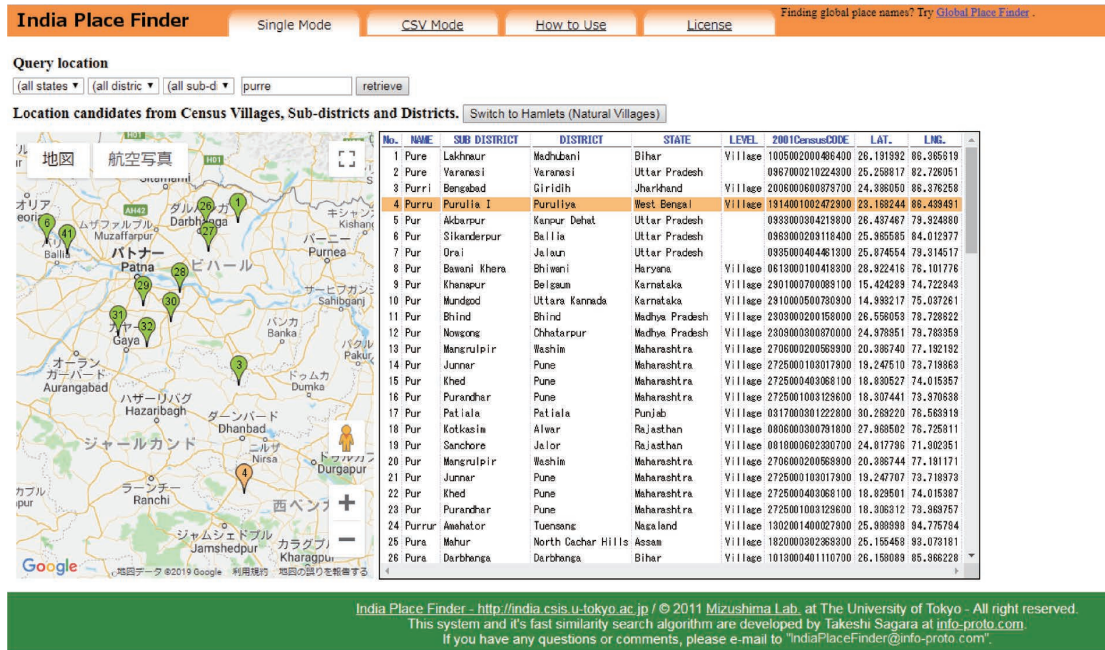
Figure 2. India Place Finder
Source: The website of India Place Finder (http://india.csis.u-tokyo.ac.jp/).

Table 1. List of cities where the locality names are available

State	City
Andhara Pradesh	Visakhapatnam, Guntur, Nellore, Chittoor, Vijaywada
Gujarat	Surat
Karnataka	Davanagere, Gulbarga, Belgaum, Hubli-Dharwad, Mangalore, Mysore
Kerala	Ernakulam, Kochi, Kozhikode, Thiruvananthapuram
Maharashtra	Mumbai (including Thane)
Puducherry	Puducherry
Tamil Nadu	Chennai, Kanchipurm, Madurai, Dindigul, Erode, Salem, Thanjavur, Tiruppur, Vellore, Perambalur, Pudukkottai, Cuddalore, Thiruvannamalai, Karur, Viluppuram, Tiruvallur, Namakkal, Ariyalur, Thoothukudi, Tirunelveli, Ramanathapuram, Sivaganga, Teni, Udagamandalam, Nagapattinam, Virudunagar, Nagercoil, Krishnagiri, Ramanathapuram
Telangana	Hyderbad

Source: made by author.

bining the list of candidates and the knowledge of users, it is possible to improve the accuracy of location identification. Users can search any place name from the roughly 600,000 villages recorded in the 2001 Census of India and from the roughly 6,000 locality names in "census" mode or from the roughly 900,000 hamlets (natural villages) in "hamlet" mode. Locality names are available for a search within the 48 cities in south India (Table 1). Users can also sort them and narrow them down by state, district, or sub-district.

In addition, if a user wants to process many place names at once, he or she can create a CSV file with a spreadsheet program, such as Microsoft Excel, and upload it instead of entering a place name in a text box one by one.

IV. Method of Making GIS Datasets from Historical Materials: An Example of Ongoing Historical GIS Study

This section explains how to make historical materials into GIS, referring to our ongoing historical GIS research project. Emphasis is placed on converting historical materials into GIS, and spatial analysis using GIS is not mentioned.

This study intends to clarify the development of joint stock companies in India during the early 20th century. The primary historical material used was the List of Joint Stock Companies in British India. This is a list containing information on companies registered as joint stock companies and was published almost every year from 1906 to 1952. TINDAS owns all its digital image data and has compiled a database of 180,000 out of a cumulative total of several hundred thousand lists. The information included in the registration was the company's name, classification of industry (three levels of classification), year of registration, capital (authorized, subscribed, and paid), and address of headquarters. This dataset was converted into GIS using the following procedure.

1. Step 1. Reorganize data according to the purpose and the resolution of location information

In the case of historical materials, the resolution of

Record ID	Company Name	Address					Record ID	Latitude	Longitude
A	A Mills	10 CANNING STREET CALCUTTA BENGAL		CALCUTTA BENGAL					
B	B Bank	AHMEDABAD BOMBAY		AHMEDABAD BOMBAY		CALCUTTA BENGAL	A, C, D	22.56552	88.36021
C	C Coal Company	5 COUNCIL HOUSE STREET CALCUTTA BENGAL		CALCUTTA BENGAL		AHMEDABAD BOMBAY	B	23.02456	72.57419
D	D Fund	1 CLIVE STREET CALCUTTA BENGAL		CALCUTTA BENGAL		AMRITSAR PUNJAB	E	31.62591	74.87257
E	E Electric Supply	AMRITSAR PUNJAB		AMRITSAR PUNJAB					
		Uniform the resolution		Reorganize the dataset		Assign the geographical information			

Figure 3.   Example of the procedure of step 1 and 2 in our study
Source: made by author.
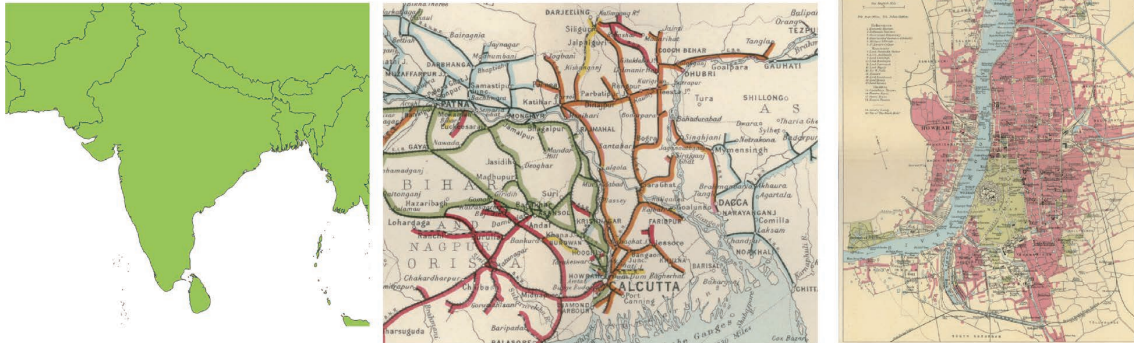
Figure 4.   Examples of base map used in our study: Coastal line with current national boundaries (left), Railways of Bengal Presidency in 1931 (center), and City map of Calcutta in 1931 (right)
Sources: Shapefiles from GADM merged by author, A part of Map of Railways & Inland Navigation in The Imperial Gazetteer of India, vol. XXVI, Atlas 1931 edition, Map of Calcutta in The Imperial Gazetteer of India, vol. XXVI Atlas 1931, respectively.

the location information may vary even in one dataset.[7] When information with different resolutions is expressed in one map, it can be misleading; therefore it is necessary to make efforts to make the resolution of data uniform according to the theme of the map (Figure 3). In this case, the company's location is described as follows:

"10 CANNING STREET, CALCUTTA, BENGAL." (A)
"AHMEDABAD, BOMBAY." (B)

While A represents the house number and the street name in Calcutta City, B represents Ahmedabad City (or Ahmedabad District) in Bombay presidency. In this example, if the map intends to provide the distribution at the city (or district) level for all of India, the data are aggregated by city using "Calcutta" and "Ahmedabad" as the location information. Conversely, if we want to see the distribution in Calcutta city, we need to extract the locations of other businesses in Calcutta, identify detailed locations like "10 Canning Street," and reorganize the dataset.

2.   Step 2. Covert place names (or addresses) to location information for GIS

The simplest and most common approach to creating a dataset that can be used in GIS is assigning latitude and longitude to place names. In this study, the place names and addresses were converted to points by assigning lati-

tude and longitude for all the resolutions of the maps. To identify latitude and longitude based on a place name, a search using a web map service such as Google Maps, a place-name database such as India Place Finder (mentioned above), or the database created by the National Geospatial Intelligence Agency[8] can be used. When a dataset of a very narrow range, such as the location information of houses in a village, is converted into a GIS, methods such as carrying a GPS at the time of fieldwork, identifying the location later using an aerial photograph or the like, and georeferencing the village map that the researcher acquires or creates can be helpful. When data from the same administrative unit, such as state or district, are fully available, it can be connected to the polygonized map by giving the Census ID. In this study, we used India Place Finder and a Google map search to identify place names in cities.

3.   Step 3. Select a base map

It is necessary to choose an appropriate base map according to the purpose of the map. Georeferenced historical maps can be base maps if they are available. If the theme of the map is related to geographic features such as altitude, climate, or soil classification, a topomap may be suitable. In this study, we used a map of only the coastline as a base map to show the district-level distribution of

companies in the whole of India, a historical map with railway lines of the days and bodies of water (rivers and canals) to show their distribution in one presidency (or state), and a historical city map to show their distribution in the city (Figure 4).

4. Step 4. Create maps

GIS software is used to create maps. One advantage of presenting historical material on a map is that new hypotheses are often intuitively derived from the map and can be expressed clearly on the map. The logic can be deepened by setting and testing hypotheses repeatedly on a map. However, maps made from incomplete information can be misleading. There is a partial lack of information in historical materials for various reasons. When they are mapped, it is difficult to determine whether information is missing, or data are zero. It is necessary to carry out the GIS analysis while recognizing these risks.

This is the procedure to convert historical materials into the GIS used in this study. This is only an example, and appropriate and efficient methods should be adopted to suit the nature of the material and the purpose of the map.

V. Conclusion

In this paper, historical GIS-related materials possessed by TINDAS were introduced. The catalog of historical materials in the collection, including those not introduced here, is available on the TINDAS website. We are currently enhancing historical materials, including GIS materials, so please check our website for the latest information. In addition to updating India Place Finder, TINDAS is also planning to release a new system for searching Asian historical maps. We hope many researchers will use these historical materials and databases to further advance their research.

Notes

1. The Center for South Asian Studies, Institute for Advanced Global Studies, Graduate School of Arts and Sciences, The University of Tokyo.
2. http://www.tindas.c.u-tokyo.ac.jp/
3. https://gadm.org/
4. http://india.csis.u-tokyo.ac.jp/
5. If the boundary of village is not fixed, only the point is shown.
6. For more information on the use of CORONA satellite imagery in India, see Kumahara (2008).
7. Current statistics or datasets collected from a field work are less likely to cause this type of problem.
8. https://www.nga.mil/

References

Kumahara, Y. (2008): Preliminary topographical analysis using CORONA satellite photographs in case of the western India. Osada, T. (ed.): *Environmental Change and The INDUS Civilization Project Report 2007*, Research Institute for Humanity and Nature, Kyoto, 41–49. (in Japanese)