CHINA HISTORICAL GEOGRAPHICAL INFORATION SYSTEMS PROJECT

DRAFT DATABASE DESIGN AND GEOCODING SYSTEM

(Version 011/09/00, Copyright)

Lawrence W. Crissman
Australian Centre of the Asian Spatial Information and Analysis Network
School of Asian and International Studies
Griffith University
Nathan, Q 4111, Australia

This document outlines proposed database design and geocoding principles for unifying GIS work on China's historical administrative geography to be undertaken by the following institutions and persons (as well as possible others):

Academia Sinica
  Fan I-Chun, Institute of History and Philology
  Eric Yen, Computing Centre

Fudan University
  Ge Jianxiong and associates, Research Center and Institute of Chinese Historical Geography

Harvard University
  Peter Bol, Department of History
  Merrick Berman, engaged by Harvard Yenching Institute

Griffith University
  Lawrence Crissman, Australian Centre of the Asian Spatial Information and Analysis Network (ACASIAN)

Background Discussion

The aim of this document is to describe the principles of a flexible, open-ended temporal-spatial database design that will not prescribe the sources or nature of the information that can be included or how China's historical administrative geography will be graphically represented. It explicitly allows for versioning of both attribute and spatial data, particularly in terms of information sources and base maps.

During the 1980s, the Cartographic Publishing House, with sponsorship from the Chinese Academy of Social Sciences, published the Historical Atlas of China, an eight volume set produced under the direction of Chief Editor Tan Qixiang. Ge Jianxiong, Director, Institute of Chinese Historical Geography, Fudan University, and his group have been working on revising Tan's work, and are prepared to produce more detailed new materials, particularly for the Ming and earlier periods, if sufficient funding can be made available. They are interested having electronic versions of their work incorporated into Geographical Information System (GIS) databases for China's historical administrative geography.

The Academia Sinica Computing Centre, under urging from Fan I-Chun and under the direction of Eric Yen, has been working successfully on digitising Tan's Historical Atlas of China, and have worked on historical administrative codes (referred to as AS codes herein). They are interested in using Taiwan's extensive historical resources to produce more accurate digital representations of the Qing and Republican Period administrative systems, and are prepared to cooperate with Fudan to produce a unified GIS for China's historical administrative geography.

In the mid-1990s, the late Robert Hartwell, University of Pennsylvania, began work in conjunction with the China in Time and Space (CITAS) Project to co-locate historical administrative units down to the county (xian) level with their modern counterparts, using spatial data for 1992 counties in MapInfo format produced by L. Crissman at ACASIAN. At the time of his death in 1996, Hartwell had in retirement completed work on Tang, Northern and Southern Sung, and Ming data/maps, and was well along on work on coding of the Qing and Republican Period administrative systems prepratory to generating maps for 1820 and 1911. His property, both intellectual and real, was bequethed to the

Harvard Yenching Institute, which has only recently taken possession. Peter Bol is committed to making Hartwell's China Historical Studies data available for scholarly use, and is supportive of efforts to produce more accurate representations of China's historical administrative hierarchies. He is prepared to seek funding to support work on producing a spatio-temporal GIS for China's historical administrative geography. Merrick Berman has recently been engaged to work on the Hartwell historical maps and attempts to obtain funding for the Chinese Historical GIS Project.

Since 1980, the PRC has used an official Guobiao system for coding administrative divisions (referred to herein as GB codes), which has been published by Guojia Jishu Jianduju (National Technical Supervisory Bureau) under the English title Codes for the administrative divisions of the People's Republic of China with the identifying code GB 2260 – plus a year. (The 1995 and 1999 editions have the code GB/T 2260 – and the year. Other years of publication were 1980, 1982, 1984, 1986, 1988, and 1991, the latter being current to 01/01/1992.)

Spatial datasets for China's contemporary county-level administrative geography are currently available from a number of sources. ACASIAN has produced a number of such county-level datasets since the early 1990s. One series was compiled from the 1990 Land-use Map of China and many other more recent sources, and covers the period from 1980 when the first edition of GB 2260 was published through 01/07/99 the date the information in GB/T 2260 – 1999 is current to. These ACASIAN county and urban district data are updated yearly on the publication of new editions of Zhonghua Renmin Gongheguo Xingjengqu Huajiance (Administrative Handbook of the PRC), and include shixiaqu (urban districts) for many large cities for which suitable maps have been obtained. G. W. Skinner has cooperated in refining these data, particularly for 1990-1992. Another set of ACASIAN administrative polygons were produced from the 1988 edition of the Atlas of Chinese Postal Codes under a copyright agreement with the Harbin Cartographic Publishing House. They are current to 01/01/1992, the currency date of GB 2260 – 1991, and were the basis for Hartwell's co-location of historical administrative units. Liu Chuang, at CIESIN and on behalf of the CITAS Project, produced two time variants of county-level polygons, one for the census date of 01/07/1990 and one for 31/12/1990, from materials provided by the Chinese Academy of Surveying and Mapping. They are available from CIESIN's SEDAC China Dimensions Data Collection and from the China Data Center, U. Michigan. ArcChina, produced by the National Bureau of Surveying and Mapping and distributed by the Environmental Systems Research Institute (ESRI) includes county-level polygons current to 1990.

In order to accommodate the continual changes to China's contemporary administrative geography, ACASIAN designed a spatio-temporal database system that will extract spatial data for any date covered by their databases and create a time specific GIS coverage. A 1992 paper by Paul Yates and L. Crissman (http://www.asian.gu.edu.au/papers/historical.html) describes those database design principles, some of which have been adopted in this document. The CITAS WWW site contains a page (http://citas.csde.washington.edu/design/dbdesign) with a document titled 'Construction of the County Boundary Database of China' dated 27/12/1994 (last updated 12/05/1997) that describes a spatio-temporal spatial data retrieval system that is basically similar in many of its principles of operation to the one described in the 1992 Yates and Crissman paper.

The CITAS Project, University of Washington, Seattle, produced a *GB Codes of China* (1982-1994) database using the GB 2260 code books, but which records from and to dates for each code that existed during that period. It is available by ftp from the SEDAC China Dimensions Data Collection at CIESIN (http://www.ciesin.org), under the title 'GB Codes for the Administrative Divisions of the People's Republic of China'. Although new codes have been added for administrative divisions created or modified from 01/01/1992 to 31/12/1994 and some other cases, these data are largely based on GB 2260 – 1991 and earlier editions back to GB 2260 – 1982. ACASIAN has a similar database for 1980 to 1999 maintained in ORACLE that contains all administrative changes whether or not a code change was involved and additionally records the administrative units that comprise all or some of the territory of immediately prior units subsequent to any change, thus accommodating the one-to-many and many-to-one relationships involved in tracking temporal sequences of administrative control over particular places.

These kinds of temporally searchable administrative change databases, and their extension backwards in time to historical periods and dynasties, are the key elements in this spatio-temporal database design. The administrative units that existed at any point in time can be selected from them, and their GIS representations can then be assembled from components of spatial datasets. That involves using Administrative Codes, such as the GB codes and their historical derivatives, as the basis for assigning geocodes (or as the geocodes themselves) used to link to selected versions of both attribute and spatial data, as described below. As a pilot project, this spatio-temporal database design will be implemented with the ACASIAN composite data set extending from 1980 to 2000 in order to test its concepts and operations, and the results will be made available to this project.

One fundamentally important principle for the representation of historical administrative units with GIS spatial data is that they should be constructed from spatial data elements that are identical to those used for contemporary or at least later administrative units whenever there is historical or graphical evidence that they were in fact the same. In other words one should, insofar as is possible, work backwards in time when assembling historical spatial data, using any appropriate later boundary data when representing earlier administrative units. Otherwise, GIS operations such as overlays become a nightmare of sliver polygons and voids. However, that does not mean that the historical representations of China's historical field administration need to be based on any *particular* set of contemporary boundary data, such as ACASIAN datasets, the CIESIN/CITAS polygons, or ArcChina counties. This database design will allow the same dynastic administrative hierarchies and units to be represented by any sets of spatial that have been assembled and coded in accordance with its principles and procedures, despite the original base map materials they utilize.

Different GIS software systems have different spatial and attribute data structures that would make the optimum spatio-temporal database implementation for one difficult or inefficient to implement for another. Fortunately, it is possible to devise a general but flexible design that will allow versions of the database most suitable for particular GIS environments to be established as required.

Technical Discusion

Acronym Glossary

AC – Administrative Code
ACT – Administrative Change Table
AFT – Administrative Feature Table
A/I 7 – ARC/INFO 7
AS Codes – Academia Sinica historical administrative codes
AV – ArcView
GB Codes – Guobiao (National Standard) PRC administrative division codes
CHGIS database design – This Chinese Historical Geographical Information System database design
CHS Codes – Hartwell's Chinese Historical Studies administrative codes
GIS – Geographical Information System
MI – MapInfo
PRC – People's Republic of China
RDBMS – Relational Data Base Management System
SQL – Structured Query Language
SDT – Spatial Data Table
STDB – Spatio-Temporal Data Base
STV Geocode – Spatio Temporal Version/variation Geocode
STVA – Spatio Temporal Version/variation Geocode for area or polygon features
STVL – Spatio Temporal Version/variation Geocode for line features
STVP – Spatio Temporal Version/variation Geocode for point features
TAR – Tibetan Autonomous Region
TIS – Temporal Instance Suffix
TST – Temporal Sequence Table
VCT – Version/variant Code Table

STV and Extended STV Geocode Suffix Conventions

_ (underline) followed by numeral for Temporal Instance Suffixes (TISs)
, (comma) followed by alphanumeric code for versions/variations, following TISs
. (dot or period) followed by alphanumeric for point features and in STVP fields
- (dash) followed by alphanumeric for line features and in STVL fields
+ (plus sign) followed by alphanumeric for polygon features and in STVA fields

Types of Tables

Although they can be variously combined under some circumstances, as discussed below, the basic operation of this database design is best understood in terms of four separate kinds of relational data tables: Administrative Change Tables (ACTs), Temporal Sequence Tables (TSTs), Administrative Feature Tables (AFTs), and Spatial Data Tables (SDTs). As the logic and operation of these different kinds of tables is not always the same, they will be described separately in the first instance even though they can be variously aggregated. A variety of RDBMSs based on SQL (Structured Query Language) can be used for constructing and querying these kinds of tables, including even those used in MapInfo and ArcView.

ACTs (Administrative Change Tables) utilise a ‘row versioning' method in which different manifestations of the ‘same' administrative unit can be represented by multiple rows in the one database, once for each instance of the unit before and after each change it underwent or for different versions or variations of the unit as represented in different sources, etc. Any change in a unit, such as a boundary adjustment, a new capital seat, a name change, a new status or change in administrative hierarchy position, or a new Administrative Code (AC), creates a new instance of the unit and requires a new row. Since an AC, such as a GB code, may not uniquely identify such instances, they will be given unique Spatio-Temporal-Version/variant-Geocodes (STV Geocodes) that will be derived from Administrative Codes as described below that also serve to geo-reference administrative units and their boundaries and capitals/seats by identifying appropriate spatial representations in SDTs.

The most rudimentary form of an ACT needs to contain only three fields, one for the STV Geocodes that uniquely identify and link rows to the other kinds of tables, one for the date (as defined below) when a particular instance of an administrative division became current, and the other for the date when it was changed in one way or another or ceased to exist. However, ACTs will normally be more elaborate, having additional fields for other identifiers such as actual names to aid the humans who create and refer to them in addition to the ACs used in deriving their STV Geocodes as an additional basis for straightforward selections. They will also contain variant/variety field(s) if such features are included in linked spatial database(s), as well as fields that identify variant feature and spatial datasets, as described below. However, whatever else it may contain or be amalgamated with, the minimal components of an ACT can be queried to select all and only those instances of administrative divisions that were in existence on any given date, and the selection can be matched with rows in the other kinds of tables using the STV Geocodes.

Although it would be possible to have a single, unified Administrative Change Table (ACT) for all versions of all periods and dynasties, it is no doubt more practical to divide the overall Chinese Historical Administrative Geography database into separate sets of ACT tables for each period/dynasty, and perhaps even for each major version of the administrative system within a period/dynasty (if there are any such).

TSTs (Temporal Sequence Tables) allow before and after linkages among instances of administrative units, also utilizing a row versioning system that differentiates among the possible multiple precursor or successor instances or units that involve many to one or one to many relationships. There will be a row version of each precursor admin unit for each of its successor units that occupied some or all of the territory. Selecting for a unit will therefore produce its set of successor units, and selecting a successor will also yield all of its precursor units that have it in one or more of their row versions. There will also be unique STV Geocodes for each row. Spatial relations between precursors and successors can be described in a ‘notes' field, but are best represented in a GIS environment by the linked spatial objects selected from SDTs.

AFTs (Administrative Feature Tables) contain attribute information for the administrative units, such as their names in multiple orthographies, the names of their capitals or seats, and their level in the administrative hierarchy at the time, etc. The Identical STV Geocode fields in 'linked' ACTs and AFTs will allow them to be related or joined, producing a more elaborated set of attributes for administrative units selected from an ACT. A single ACT may be linked to more than one AFT, such as ones representing information in different orthographies or languages or from different sources, requiring 'Feature Dataset' fields in the ACT to identify the particular AFT to be linked in any instance.

SDTs (Spatial Data Tables) operate in much the same manner as ACTs and their 'linkable' AFTs, as each contain the same identical STV Geocode fields that link administrative units in an ACT to their spatial data representations held in or accessible by the SDTs. As different SDTs can be linked to the same ACT, the latter also require 'Spatial Dataset' fields that identify the SDTs that contain spatial data of various kinds (different basemaps, high or low resolution; region/polygon elements, boundary lines, capital points, etc.) that can be used to assemble the spatial objects needed to represent the administrative units selected from an ACT.

Combination Tables

As shown below, there can be advantages in ease of selection and representation of the information in ACTs, AFTs, and even TSTs if their fields are joined permanently to SDTs, thus possibly eliminating the need for Geocodes (see below) apart from their use as a basis for identifying, at least initially, the spatial objects chosen to represent an administrative system. However, no matter how they may have their fields variously combined, temporarily or permanently, in order to facilitate certain operations of this database design, there are advantages to creating and maintaining all four kinds as separate RDBMS tables.

Administrative Codes and Geocodes

Administrative Codes (ACs) are simply artificial or arbitrary alphanumeric 'names' that identify administrative entities. Their principal advantage over the ordinary names of administrative units or capitals/seats are that they uniquely identify elements in an administrative system, as no two are the same, and they can also embody the hierarchical relationships in the administrative system. ACs can also be used as Geocodes if they additionally identify spatial objects representing administrative units in a geo-referenced spatial database, e.g., one whose coordinate system relates to the earth's surface. The Spatio Temporal Version Geocodes (STV Geocodes) employed in this database design will be derived from agreed-upon systems of ACs for China's historical administrative geography, whether modeled on GB codes, AC codes, or Hartwell's CHS codes, or some other system. They will also incorporate additional suffixes that will uniquely identify temporal instances and versions/variants of various kinds of administrative spatial objects, and can therefore be used to link them to their attributes held in AFTs and TSTs, and can also identify and/or select their visual and topological representations from SDTs.

The official PRC six-digit Guobiao administrative division codes (GB codes) discussed above are an excellent example of an hierarchical administrative coding system, and make a very useful model for devising the ACs required for this Chinese Historical GIS database design. Insofar as they go, the GB Codes represent the hierarchical structure of the People's Republic of China's contemporary administrative system of provinces and provincial-level municipalities and their equivalents such as the Tibetan Autonomous Region or TAR (the first two digits in a GB code), which are divided into prefectures and prefectural-level municipalities and their equivalents (second set of two digits), which in turn are divided into counties and county-level municipalities and their equivalents (last two digits). The last two digits are also used under certain conventions to represent urban districts in large municipalities. When the last two digits are naughts (zeros), the administrative unit is at prefectural-level, and when the last four digits are naughts (zeros), the administrative unit is at provincial-level. There are also various conventions for assigning the last two digits, but they have not always been strictly followed.

During historical dynasties when the Chinese empire had more than three levels in its administrative hierarchy down to the basic county-level, additional sets of two digits are needed, either at the beginning of the set for such things as supra-provincial 'circuits', or sometimes between the provinces and (usually) the prefectures, or even both. Hartwell's CHS historical coding system employed ten digit codes for this reason. It will also be very useful to distinguish code sets that pertain to particular periods/dynasties with unique letter prefixes, as Hartwell did. However, when such extra digits are employed in forming administrative coding systems, the last two uninterrupted (or 'non-suffixed') digits will still always be used for the county (or contemporary urban district) level(s) of the administrative hierarchy, as counties (*xian*) and their equivalents have always been the basic building blocks of Chinese field administration. If sub-county administrative units, such as townships (*xiang* and *zhen*), are to be included, they can be identified with alphanumeric suffixes that are preceded by a plus sign (+), indicating polygon features.

From and To dates

Different temporal instances of an administrative unit will have distinct STV Geocodes as described above. However, in order to select such different temporal instances of an administrative unit, the date when a unit was established or when a change was made need to be included in ACT fields representing the 'from' and 'to' dates when the particular instance was current. In a 'Current From' field, a date known by its year only or its year and month only will be treated as the first day of the year or month, whereas in the 'Current To' field such incomplete dates will be treated as the last day of the relevant period. Such dates in the current era (CE) will be recorded in a yyyymmdd format if a precise date is known, in a yyyymm00 or yyyymm31 format if only a month and a year are known, and with yyyy0000 or yyyy1231 if the year alone is all that is known. If the only information about the establishment, modification or dissolution of an administrative unit is a reign period, or even a dynasty, then its first year will be recorded as the 'from' date and the last year as the 'to' date. BCE dates will have a leading minus (-). For instance, if a name change was known to have occurred some time in CE 976, the Current From date of the new record would be 9760000, but the Current To date of the previous instance or precursor(s) of the administrative unit would be 9751231. However, if a change occurred in BCE 124, the Current From date of the new record would be -1240000, but the Current To date of the previous instance of the administrative unit would be -1251231. Querying the database will result in the selection of all the instances of administrative units that were in existence on a particular day, month, or year provided that there were no changes within the selected month or year, in which cases more precise dates would be required to produce a valid selection. Alternatively, there could be multiple from and to date fields for years only, years and months, and years, months, and days. These principles may require different specific implementations in different RDBMSs that store dates in proprietary formats.

Temporal Instances

Because an administrative unit can change in various ways without necessarily being assigned a new GB code or other kind of AC, unelaborated ACs may not uniquely identify different instances of the same admin unit and cannot represent different versions or variants, without some elaboration. Hence the need for the Spatio-Temporal-Version/variant-Geocodes (STV Geocodes) that will uniquely identify instances as well as versions/variants of the 'same' admin unit in order to link to AFTs, SDTs and TSTs. Version/variant conventions for producing STV Geocodes from ACs are described below. Here it is only necessary to establish the convention(s) to be used to distinguish different temporal instances of admin units that have the same ACs. That will be accomplished by adding numerals in a temporal progression beginning with '1' following an underline (_) symbol to indicate the first in a temporal series of instances of the 'same' admin unit that has the same AC. Note that since instances of the same administrative unit are distinct in time, selection of instances from the same ACT would preclude having more than one such instance in any selection set.

As an example of temporal instance suffixes for Administrative Codes, which makes them a form of STV Geocodes, the number of Beijing's urban districts (*shixiaqu*), which together have a GB code that is equivalent to prefectural-level, were increased on 29/04/1997 by the inclusion of erstwhile Tong Xian to the east of the previous group as Tongzhou Qu, although their GB inclusive code, 110100, was not changed. Since that was the first change to Beijing's urban district set since the GB codes were established in 1980 (apart from a change in the urban district seat (*shixiaqu chengguan*) of one constituent district on 17/11/1997, which was not a change in prefectural-level set of districts, per se), the original set will be coded with 110100_1 and the enlarged one with 110100_2. 110100_3 will be used for the new set formed on 03/03/1998 with the inclusion of erstwhile Shunyi Xian as Shunyi Qu. As they are temporally distinct, only one of those instances of 110100 could be present in any selection set derived from an ACT.

Versions and Variations

Historical sources may not agree on how China's administrative geography was hierarchically organized at particular points or periods in the past. To use a current example, contemporary maps of China produced in the ROC are quite different even at the provincial level to those produced in the PRC. Therefore, different versions of the database would need to be maintained to represent the two distinct hierarchical organizations of China's contemporary administrative system if both were to be included in the database. This can be accomplished by having an ACT 'Version' field coded with some acronym or other mnemonic referring to the government or source(s) on which the alternative 'Admin Code' and derivative STV Geocode assignments are based. Variations can also exist within a version. For instance, as far as the central authorities in Beijing are concerned, four new counties were established in the TAR in 1988, but apparently no one in the TAR itself ever implemented those decisions. The result is that some maps of the TAR published in Beijing are not the same as those often produced locally, which actually represent all and only those counties for which any statistical data, including the 1990 census returns, are available. Also, different maps can show different boundaries for the same instance of an administrative unit, and such minor variations are also not uncommon for TAR counties, or even prefectures.

Differences in a version code could be included in an ACT Version Field to identify the two (or more) alternatives for instances of the same administrative units contained in different rows of the ACT database. For instance, the Version of China's contemporary administrative hierarchy based on the PRC GB codes could have 'P' in the Version Field for cases in which there were no PRC alternative versions, but could have such things as P,XZ1 and P,XZ2 to distinguish variations in the TAR, or could even have P,XZ2,1 and P,XZ2,2 to distinguish minor alternative boundaries of the basic XZ2 variation. (These examples use the standard ACASIAN two letter codes for contemporary PRC province-level units, such as XZ for Xizang, as contained in Appendix One.) Alternatively, 'Variation' field(s) could be included, either in addition to, or in case the ACT was limited to a single version, in place of the Version field, in which case the terms separated by commas in the examples would be in two or three separate fields, the latter examples requiring a second 'Minor Variation' field. In any event, there either need to be Version, Version/Variation, or Variation fields in any ACT which incorporates such differences in administrative entities, or there would need to be a separate Version/variation Code Table (VCT) for each ACT for which versions and/or variations existed. Or both, as desired.

<u>STV Geocodes</u>

Additionally, version/variation indicators need to be suffixed to Administrative Codes to create STV Geocodes after any instance indicators and following a comma (,) (or some other diacritical sign depending on what the RDBMS allowed). For instance, the versions/variations illustrated above could be represented by the following suffixes: ,PXZ1 and ,PXZ2 or even ,PXZ2,1 and ,PXZ2,2. Each such suffixed STV Geocode would need to be a row version of the basic STV Geocode, or they could be stored as row versions in the separate VCT (Version/variant Code Table). The reason for having such version/variation specific Geocodes is that they would be used to select the appropriate spatial objects from SDTs that contained or could access the needed version/variant spatial objects. Doing a second selection on a ACT, or an additional one on a related VCT, and sending unique results to an SDT is a better solution, I think, than selecting for versions/variations in an SDT itself, although that is also an option. Some means are necessary to ensure that only one spatially distinct version/variation could be included in any one selection of STV Geocodes. That should be readily accomplished, as they would have the same AC and more than one occurence of the same AC in a spatial object selection could be precluded, apart from instances of administrative units that need to be represented with multiple polygons, such as ones that include islands or exclaves, which could be identified in a separate field. The STV Geocode for an administrative unit that had neither instances nor versions/variants would simply be its Administrative Code without suffixes following an underline (_) or comma (,).

<u>Operation of Spatial Data Tables (SDTs) and Extended STV Geocodes</u>

STV Geocodes, including those with possible suffixes as required to uniquely identify different instances and/or versions/variations of their administrative units, can be selected from ACTs (or ACTs and VCTs) in terms of when they were current, and can in turn be used to identify successor or precursor units from TSTs (Temporal Sequence Tables), to merge additional attributes from AFTs (Administrative Feature Tables), and to select geo-referenced spatial representations from SDTs (Spatial Data Tables). Temporal Dataset (TD), Feature Dataset (FD) and Spatial Dataset (SD) fields in an ACT would be desirable if more than one TST, AFT or SDT could be accessed with the same selection of STV Geocodes from a particular ACT. Relating STV Geocodes with TSTs and AFTs is relatively straightforward, but things are more complex with regard to SDTs because the same administrative units, or instances or versions/variants thereof, can be represented in a GIS by either points, lines, or polygons, which can correspond respectively to their capital/seat locations, their boundaries, or their territorial areas (or 'footprints').

Indeed, as discussed more completely below, in some GIS environments or situations it may be desirable to build polygons from selected identifying points ('labels' in ARC/INFO terminology) and their bounding lines (or 'arcs'), rather constructing them from merged polygon components. Other GIS environments lend themselves best to storing whole, perhaps partially overlapping polygons representing spatially differing instances, versions/variations, and precursor/successor administrative units themselves, in what is the spatial equivalent of a row versioning table in a RDBMS. Indeed, some RDBMS software, such as Oracle, now provide for storing spatial objects, including polygons, in a relational table itself. Since this database design does not prescribe what GIS software or RDBMS will be employed to implement the China's historical administrative geography project, a generic SDT design is necessary although specific sub-types optimized for particular GIS software environments will also be described below.

Basically, what a generic SDT will do is to access or 'load' a set of spatial objects selected on the basis of their of STV Geocodes, themselves usually selected from an ACT on the basis of a particular date when they were current and what versions/variations are desired. To be generic, a SDT should be able to access all three kinds of spatial objects, points, lines, and polygons, although SDTs designed to operate efficiently or effectively in particular GIS software environments might be limited to accessing only one or possibly two of the basic types. Therefore, SDTs will need to contain fields differentiating or at least identifying the kinds of spatial objects that it can access (or that it actually contains, in some cases). Again, this will be done with row versioning, so in a SDT each STV Geocode will normally have as many rows as it has the most components in any of the three categories, each of which would have a separate field. For instance, a county-level administrative unit that consisted of a portion of the mainland plus two islands (or exclaves, for that matter), would need as many rows as it had boundary elements, including all of the coastlines. Three (or more) of those rows could be used simultaneously for the unit's polygon elements, and one of those in turn for the capital/seat point location. (In the ARC/INFO 7 environment, each polygon would need an associated label point, but one of them could be the capital/seat location if they were identified in a separate field or by some feature of its code – see below). When a set of STV Geocodes was used to select required spatial representations of an administrative unit, a second selection could then be made in terms of the fields identifying the desired kind(s) of spatial objects.

As additional STV Geocode suffixes may be necessary to identify and select the different kinds of spatial objects that can be used to represent an administrative unit, some additional conventions are required to differentiate those suffix codes. It is proposed that alphanumeric identifiers for point objects follow a dot or period (.), those for line objects follow a dash (-), and those for polygon objects follow a plus sign (+). In the case of the administrative division discussed above that was composed of a portion of the mainland and two additional islands, comprising, let us say, ten line segments in total for the coastlines and mainland boundaries, it would have a total of ten row versions in an SDT that had, in addition to a field for the STV Geocode, three fields identifying the existence of point, line and polygon objects and three more containing Extended STV Geocodes for each type of object. The latter would contain STVP (for point data), STVL (for lines) and STVA (for areas, or polygons) Geocodes, each consisting of the same base STV Geocode suffixed with an alphanumeric identifier for the specific spatial objects following the appropriate indicator, e.g., a dot, dash, or plus sign. The ten row versions needed to identity the ten line segments can also be used as required for the Extended STV Geocodes for the other kinds of spatial representations, as they are in different fields. Alternatively, separate point, line and/or polygon SDTs could be maintained.

An additional field identifying those points that are capital/seat locations would aid in their straightforward selection, but they could also be differentiated with the use of .C, contrasting with possible .1 to .*n* or .A1 to .A*n* (for additional) for the other label points needed in ARC/INFO 7. Correspondingly, the ten line objects outlining (and necessary to construct using ARC/INFO 7) the administrative unit could be coded with -C1 to -C*n* for coastal elements and -B1 to -B*n* for boundary lines in addition to having a field that differentiates them. Similarly polygon elements, including any fragment components necessary to construct instances or version/variations, could be identified as islands, +I1 and +I*n*, as opposed to +M1 to +M*n* for mainland areas and such things as +E1 to +E*n* for any exclaves, etc. The extended STVP, STVL and STVA Geocodes with those kinds of suffixes will uniquely identify the sets of spatial objects necessary for representing an administrative unit in terms of points, lines, or polygons (or a combination of points and lines from which polygons can be built in A/I 7). A purist might argue that arbitrary sequential numbers or letters should be used for such suffixes as the fields indicating what kinds of spatial data could contain such differentiating indicators. However, I am of the strong opinion that all geocodes should contain as much mnemonic baggage as possible to aid the humans who will create, maintain, and modify such codes and the spatial elements they identify, and should also be as redundant as possible in terms of additional fields that make selections of such things as capital label points as straightforward as possible.

There is one aspect in which line features differ from points or polygons, and that is because boundary lines internal to a contiguous set of administrative units separate, and therefore 'belong' to two but only two different administrative units. This is irrelevant for those GIS software environments in which polygons are primary spatial entities and internal boundaries are duplicated (precisely if no voids are present) so each one in a duplicate pair belongs to only one adjacent polygon. However, in generic SDTs that does not matter, since one might want to represent a selection of administrative units by their boundary lines instead of as polygons, so internal boundary features need to be selectable in terms of each of the polygons of which they are components. That is most efficiently accomplished by having two STV Geocode fields, one for the first admin unit (instance or version/variant) used to generate Extended STVL Geocodes for its boundary elements, and a second one for the STV Geocode of the neighbor on the other side of a particular line element. Although one could reserve the two fields for right and left sides of boundary vectors that were made to point in a consistent direction, it actually makes no difference whatsoever which field contains which of the two possible STV Geocodes pertaining to internal boundary lines, as in any event both fields would have to be searched in combination to select all of the boundary lines needed to represent a set of administrative units, or even one such unit. Note that each boundary line would have only one Extended STVL Geocode that would allow the unique selection of its representative spatial object, but it does not matter which STV Geocode it is based on as both of the STV Geocodes belonging to the two units it divides are identifiable in the separate SDT fields. Line features on the perimeter of an administrative system can have no value, 'Nil', or 'Universal' (in ARC/INFO terminolgy) in the second STV Geocode field.

Depending on the relationship between an SDT (which is essentially an attribute table) and associated spatial data in the particular GIS environment being employed, the selection set of Extended STV Geocodes required to represent, in one way or another, the administrative units selected from an ACT would then be used to select or access the geo-referenced spatial representations required to constitute the desired coverage (in A/I parlance). Points or lines can be used without further ado to represent administrative hierarchies or boundaries, and information from AFTs can be joined to them directly on the basis of STV Geocodes. However, polygon representations will usually require one or more further steps before joining additional attributes from AFTs. (In ARC/INFO 7, the polygon coverages would have to be built, or cleaned and then built automatically, first of all.) In any GIS environment, a selection of polygons in which at least one administrative unit was composed of multiple polygons, either islands or 'fragmented' components, needs to have those components 'merged', 'dissolved', or 'regionalised' on the basis of their STV Geocodes before joining with AFT attributes.

The principles described for assigning Extended STV Geocodes can be used for features other than administrative entities. Points, such as mountain peaks or villages, can also be identified with alphanumerics following a dot (.) appended to the STV Geocode for the county-level administrative unit in which they are located. Similarly for rivers, roads, walls, etc., following dashes (-) and for area features such as lakes, mountains/mountain ranges, and other geographical areas or regions following a plus sign (+). Note that this system can be employed at higher levels of an administrative hierarchy as well by employing STV Geocodes based on prefectural or provincial Administrative Codes, depending on the extent of the objects being coded. Following the intentionally redundant coding/identifying properties of this database design, in addition to having fields that identify what kinds of point, line, and polygon features are included in an SDT, the Extended STV Geocodes should also contain alpha or alphanumeric indicators of spatial feature types as human-friendly aids.

Implementing GIS Spatial Representations in Different GIS Environments

ARC/INFO 7 Polygon Coverages

Some ARC/INFO 7 terminology and special considerations have been introduced in the previous section. It can probably be assumed that most people assembling spatial data for implementing this database design will be familiar with this GIS software as it is widely used in major GIS labs or operations at academic institutions throughout the world. It also allows for two major alternatives with respect to how polygon spatial objects can be assembled for sets of STV Geocodes. Basically, A/I 7 builds polygons from component label points and arcs (lines), and therefore all three kinds of spatial objects can exist in the same A/I 7 coverages. Although arcs can be given attributes in .aat info tables, and can be selected on that basis, the attributes of arcs have no necessary relation to those of the polygons they can be assembled into. There can be no duplicate arcs after a coverage has been ‘cleaned’, and a coverage cannot be ‘built’ (have topology generated) without being clean in that regard. All sets of closed arcs become polygons when the set of arcs in a coverage are built into polygons, and no overlapping polygons are possible. All polygon attributes (apart from the automatically calculated area and perimeter values and the arbitrary polygon # value for a polygon without a label point) are attached to their internal label points and are held in .pat info tables the same as attributes for any other point data.

Those features have disadvantages as well as advantages. First of all, there can be no voids inside or between polygons one might want to assemble out of a set of arcs, as polygons (intended ones or not but including the surrounding ‘universal’ polygon) must completely fill the ‘space’ of a coverage and no space can be occupied by more than one polygon. If a space that will become a polygon has no label point before a build, there will be none afterwards either and the polygon won’t have attribute information afterwards apart from its area, perimeter, and randomly assigned polygon #, which may have changed from the previous build. If labels are generated for polygons without label points, they will have the same info ‘items’ (equivalent to fields) as the others, but they will be empty apart from the randomly assigned # values and –ID numbers (which identify labels), as for polygons without labels apart from the latter item. If two points end up inside the same polygon during a build, A/I 7 will make their attributes identical, including their point –ID arbitrary numbers, randomly selecting one of the set of labels as the source of all attribute information for the other(s). So, if a valid set of representations of an administrative system is to be built in A/I 7, one but only one label point must be included for each intended polygon if attributes such as STV Geocodes are to be validly associated with the resulting set of polygons.

One A/I 7 advantage is that it is relatively easy to construct coherent sets of polygons as none can overlap and unintended slivers are relatively easy to identify as they will have no attributes such as STV Geocodes. Another is that since every polygon also has a label point, certain of those labels can georeference the locations of capitals/seats in the same coverage using the same basic STV Geocodes to join AFT attributes to the identified capital/seat points. Also, a coverage containing polygons can also include extraneous arcs that do not join to form additional polygons, and all arcs can have attributes such as Extended STVL Geocodes. So, one A/I coverage can contain capital/seat points, boundary lines that extend outwards into frontier areas but do not close to make polygons, and polygons for administrative units with independently selectable boundary arcs.

<u>ARC/INFO 7 Method 1</u>. One of the distinct ways in which A/I 7 can assemble polygon representations of administrative units, which was implicit in the foregoing discussion, is to build polygons from labels and arcs that have the same STV Geocodes. (This is basically the method utilized in the Historical GIS for England and Wales project.) All arcs and labels that are needed for any instance or version/variant identified in the ACT would need to be included in the coverage. Of course each arc that is part of an internal boundary will belong to two but only two polygons, and generic SDTs as well as those designed to work with this method of polygon construction should have a second STV Geocode field for codes pertaining to the polygon on the other side of each arc, as pointed out above. As A/I 7 will only select the same arc once for each selection set, no duplicate arcs can be selected from an .aat info table even if their Extended STVL Geocodes were duplicated in a SDT selection set as a result of a combined query on both STV Geocode fields. As also pointed out above, it is irrelevant which of the two STV Geocodes was the basis for the unique Extended STVL Geocode, as both basic STV Geocodes will be attributes of each internal boundary arc, although both items (fields) would need to be searched in combination to select all of the arcs forming a particular polygon’s boundaries or those needed for a selection set of polygons identified by their STV Geocodes.

Once the required arcs and label points needed to construct the polygons for a set of administrative units have been identified in an SDT, their Extended STVL and STVP Geocodes can be used to select the requisite arcs and labels from a coverage’s .aat and .pat info tables if they are coded with the relevant Extended STV Geocodes. Alternatively, a coverage’s .aat and .pat info tables could ***be*** the line and point SDTs, and their Extended STV Geocode items (fields) would be used to select the arcs and labels needed to build the polygons for a particular time and set of versions/variants once they had been copied into a new coverage. Although that might be thought to obviate the need for Extended STV Geocodes per se, the multiple arcs and even labels needed for compound polygon features would still need to be selectable with something like the Extended STV Geocode suffixes, which are not necessarily unique without the STV Geocodes they are joined to. (Randomly generated unique identifiers are not human-friendly.)

Once the component arcs and labels have been selected and put into a coverage, it needs to be built and then, if the .pat file and the SDT are not the same, the Administrative Codes (ACs) from the AFT need to be joined on the basis of the STV Geocodes onto the Extended STVP Geocodes in the .pat info file. When that has been done, regions can be dissolved on the basis of the ACs. ‘Regiondissolve’ can be used again, and again, to generate regions for higher levels in an administrative hierarchy within the same coverage once the ACs for the superordinate administrative units have been joined from an AFT. If desired, the component polygons of the regions can be eliminated from a coverage with ‘regionclean’ in permanently stored time slices and version/variant coverages. Alternatively, the new polygon features produced by such regionclean procedures could simply be added to the coverage and SDT as new polygon objects with new STV and Extended STVA Geocodes based on the Administrative Codes for their levels in the administrative hierarchy.

<u>ARC/INFO 7 Method 2</u>. The second method of assembling polygon coverages for administrative systems in A/I 7 is based on the 1992 ACASIAN paper, which employs a space-time composite as described by Langran (1991). In a space-time composite, areal features such as administrative units are decomposed into the minimum polygon elements that are identified as being part of one and only one version/variant of an administrative unit at any one point in time. Those polygon ‘components’ can be generated by overlaying polygons (or arcs and necessary labels) representing all instances and versions/variants of all administrative units in an ACT. When cleaned and built, the coverage would be fragmented into all of the necessary component polygons representing the minimal elements that combine in various ways to constitute any particular set of instances and versions/variants of administrative units selectable from an ACT.

Each of the resulting polygon components would inherit STV Geocode items (fields) from the coverages that were overlain, but the item names in the original .pat tables would need to be differentiated beforehand in some simple manner such as STV1 to STV*n* so none would be lost and all of those inherited by each polygon component could be used to create a row version in the SDT for each STV Geocode belonging to each component polygon. For purposes of unique identification and joining information from an AFT, the component polygons would be assigned an Extended STVA Geocode based on one of its STV Geocodes, but as with the Extended STVL Geocodes used in the first A/I 7 method it would not matter which one was used for that purpose, so the most ‘basic’ ones should be given preference (eg., mainland polygons as opposed to islands, etc.). The Extended STVA Geocodes would then serve as mnemonics indicating at least one prominent administrative unit that a polygon component would combine with others to form, as well as being the basis for joining information from the AFT onto each of its polygon components.

All administrative units that a particular polygon component could be a part of would have their STV Geocodes in one of the row versions belonging to the polygon component in the SDT. However, the SDT would have to be a relational table separate from the .pat info file for the coverage containing the polygon components unless it were structured with as many items (fields) containing STV Geocodes as there were any temporal changes in the ACT. That would only be a practical solution IF there were only a very limited number of times when administrative changes were known from the sources. Otherwise, selections from the SDT derived from temporal and version/variant queries on an ACT would be used to join Adminstrative Codes from an AFT to the polygon components in the .pat info file for the coverage, using the selected STV Geocodes as the common identifiers between the AFT and the Extended STVA Geocodes. The label points representing capitals/seats would need to be separately identified in like manner.

Then, since all of the component polygons needed to constitute any of the temporal alternatives, instances, and versions/variants of an administrative system are already in the same coverage, but are now coded with their selected Administrative Codes (ACs), all that is needed to represent the ACT selection in the coverage is to use the regiondissolve command on the newly joined AC item(s). This could be done at all desired levels of the administrative hierarchy manifest in the ACs, and ‘regionclean’ could be used to eliminate the polygon components and their unneeded arcs from the coverage if desired. Arcs that do not form polygons can also be included in the coverages employed in this method, but would need to be separately selected and identified, and the particular polygon component label points could be identified as captials/seats at particular times as in the first method.

An alternative means of ensuring that the correct AFT information ended up in the correct polygon after a regionclean procedure would be to select labels alone from a distinct label coverage containing all and only the label points needed for all of the administrative units (and their islands and exclaves) on the basis of their Extended STVP Geocodes matching a selection set from the ACT, join the desired AFT information, and then replace the regionclean polygon labels with the new set of label points before forming regions for different levels of the administrative hierarchy with regiondissolve. Such a label point coverage containing only capital/seat locations could be used alone to represent the administrative system in terms of point objects in either A/I 7 method.

Advantages and Disadvantages of the Two A/I 7 Methods

A/I 7 Method 1. If starting from scratch with maps showing boundaries and capital/seat points, Method 1 would be the obvious choice. Arcs representing all boundary changes for the time period covered by an ACT and all included versions/variants could be combined into the same coverage. A consistent choice would need to be made about what source(s) would be used for the standard arcs representing boundaries that did not change over time. Arcs for boundaries in frontier areas that do not join could be included, but all unwanted dangling nodes, etc., could be eliminated using standard A/I procedures. Coding of all arcs with Extended STVL Geocodes could be done as they were digitised, but care would need to be taken to ensure that all arc segments of a boundary ended up with the correct Extended STVL Geocode after the arcs in the coverage were intersected. Label points would also need to be added to the coverage, NOT for every polygon but only for those needed to compose the administrative units in the AFT, including those for islands and exclaves, etc. However, labels for the point locations of capitals/seats for the administrative units that have unclosed boundary arcs could also be included, as multiple labels ending up in the Universal polygon are allowed in A/I 7 coverages, and those points could also be used alone to represent an administrative system in terms of point objects. The coding of arcs and label points with Extended STVL and STVP Geocodes could be tedious with this method, and would need to be done with great care to ensure that all such coding was 100% accurate.

A/I Method 2. If starting with existing polygon coverages for different points in time, Method 2 would be the obvious choice, although polygon borders that did not change over the time period covered or did not belong to different versions/variations would need to be absolutely identical. Those temporal 'slices in time' polygon coverages, as well as any for versions/variants would be overlain and the polygons intersected to produce all of the polygon components that could be re-assembled in various combinations to represent the administrative system at different points in time. Arcs forming the polygon boundaries would not need to be coded with Extended STVL Geocodes at all, eliminating that tedious coding job, although arcs not forming polygon borders should still be coded and made selectable in a separate item (field).

ArcInfo 8 and ArcGIS Phase III

This year's new ArcInfo 8.0 software is thankfully completely backward compatible, so any Spatio-Temporal Database (STB) established in ARC/INFO 7 will work in ArcInfo 8. However, a number of new components have been added to ArcInfo 8 including ArcMap, which will become the same as ArcView 8.1 in next year's Phase III of ArcGIS and has some characteristics in common with ArcView 3.2. One presumes, therefore, that any STDB established in ArcView 3.X would also work within ArcInfo 8's ArcMap. I am not yet familiar enough with all of the new features of ArcInfo 8 and in particular 8.1 to confidently design an implementation of this database design that would optimize creation and operation of the system in the new environment, however there are some exciting possibilities. A/I 8 has an object oriented data model, called 'Geodatabase', in which the 'behavior' of an object can be stored along with it in ArcSDE. So, both the geometric shape of objects **and** their attributes are in the same database, which can actually reside in MS Access v. 8 at least in some cases! So, in an A/I 8.1 implementation of the CHGIS database, the SDT would contain the spatial objects that matched its Extended STV Geocodes, and ACTs, AFTs and even TST could be combined into a single RDBMS table.

ArcView and MapInfo

A basic difference between ARC/INFO 7 and both ArcView 3.2 and MapInfo Professional 6.0 (and earlier versions of either) is that only one kind of spatial object can be stored in a set of ArcView shapefiles (.shp) or MapInfo table files (.tab), so separate point, line, and polygon themes or table files will be required to implement this SDB Design if all three kinds of spatial objects are to be included as alternative representations of an administrative system. However, one major advantage of AV polygon shapefiles is that polygons need not fill all of the 'space' represented in the shape files, and polygons can in fact overlap, in whole or in part, a feature not often implemented as AV shape files are most often created from A/I coverages. Unlike A/I, AV treats polygons, as well as merged 'regions', as primitive not derivative spatial objects, and polygon borders do not exist apart from the polygons they define except in a separate AV line theme. Polygons representing administrative units consisting of more than one discrete area, such as the one discussed above that included part of a mainland plus two islands, can be merged to create a single object in the AV data table. MapInfo uses different terminology, but shares the all of the above features of data format and storage with ArcView.

The significance of those ArcView and MapInfo features for this database design is that all of the spatial changes and versions/variations in boundaries in an ACT can be stored as row versions in an AV attribute data file (.dbf) or MI data file (.dat), with each row version directly linked by the software to a matching polygon, point, or line set in the associated AV theme datafile (.shp) or MI map file (.map). So, an SDT can ***be*** the AV attribute table file for a theme or a MI table file. In fact, if the From and To Date fields are included, the ACT can also ***be*** the same AV or MI data table, and selections from it will automatically produce the matching spatial objects as a selection set that can be displayed and copied to a new AV theme or MI .tab file set. There is also no need to leave the AFT information out of the same data files, or for that matter the TST information either. A time plus version/variation set of queries on such an AV or MI data table would directly select the desired spatial representations of the administrative units along with all the kinds of attribute data, and they could then be merged or combined on the basis of their Administrative Codes for higher levels in the administrative system in separate copies of the theme or .tab file set. Point and line data tables would need to have the capitals/seats and external boundary lines of higher administrative levels identified in relevant fields for this to work, but that is true for A/I 7 implementations as well. Therefore, for either MapInfo or ArcView implementations, there would seem to be no need for separate ACT, TST, AFT, and SDT files, and therefore no apparent need for the STV Geocodes that link between them or especially for the Extended STV Geocodes that identify particular spatial objects!

So, why would anyone in their right minds implement this database design in ARC/INFO 7 rather than MapInfo or ArcView? There are actually a number of reasons, largely having to do with the limited means of creating or importing new spatial data in both software systems, polygons in particular. Apart from drawing new polygons individually, one after the other either wholly on-screen or with a digitizer, neither seems to have a means to generate polygon typology from sets of lines, at least in my experience. However, that is limited in this regard given the methods developed at ACASIAN for generating polygon spatial data from published map sources which use neither MI or AVsoftware programs, and I welcome corrective information that will show me how either AV or MI can import CAD (Computer Aided Drafting/Design) cartographic drawing or design files and create polygon typology from them. Either program will readily import ARC/INFO polygon coverages and transform them into polygon themes or area objects, but that is because the polygon topology has been generated in A/I by building arcs and label points to make polygons. Either will also import just line or point data from A/I polygon coverages, but to the best of my knowledge (again, I welcome correction) neither will covert polygons to lines and points, meaning that both kinds of themes or table file sets would have to be generated or imported and identified separately to represent administrative systems in terms of those kinds of spatial objects.

If either program could import CAD data to make polygons or areas, or even if only lines or points were imported, there would remain the problem of initial identification of the new objects. Unless all of the information equivalent to the contents of combined ACTs, AFTs, and possibly TSTs (if all STV Geocodes were left out) were imported along with the CAD data and then somehow associated with the newly generated points, lines or polygons, those new spatial data would still have to be coded in some way functionally identical to Extended STV Geocodes in order to join the SDT STV Geocode fields needed to join ACT, AFT or even TST fields. Equivalents of Extended STV Geocoding would also need to be done for point, line, and area (polygon) features separately. In addition, without the equivalent of SDT Extended STV Geocode fields attached to alternative spatial data representations, such as high or low resolution versons or those based on different base map data, all such datasets would all have to repeat the whole complement of SDT, AFT and ACT fields.

In other words, it might be advantageous to import ARC/INFO 7 coverages using this database design into either MapInfo or ArcView in order to assemble and code the overlapping line or polygon sets that work most advantageously in these two software environments in order to take advantage of their comparative ease of operation in terms of selecting and representing changes or variations to administrative systems, to say nothing of the relative ease with which these software programs can be learned. However, there would appear to be greater difficulties in implementing a spatio-temporal database with this design from scratch in either ‘desktop mapping system’, or however their vendors are now characterizing them. Note also that nothing precludes permanently joining ACT, AFT, or even

TST fields to SDTs in either A/I 7 method of implementing this database design, and that could well be done before they are imported into either AV or MI.

TimeMap View

I believe that the best way to implement this database design in TimeMap, which is being adopted by the Electronic Cultural Atlas Initiative (ECAI), would be to submit ArcView or MapInfo versions to the TimeMap Metadata Clearinghouse. Ian Johnson, Archaeology Computing Lab, the creator of TimeMap, will surely advise whether that opinion is correct.

Bibliography

Langran, Gail, Time in Geographical Information Systems. Technical Issues in Geographical Information Systems Series. London: Taylor and Francis, 1992

Additional Sections

I have plans to add sections to this document outlining the procedures for preparing spatial data for implementation of this database design.

I also intend to describe all of the fields that are necessary or desirable for ACTs, TSTs, AFTs, and SDTs

However, the principles to be employed in the design have all been spelled out and the composition and operation of the STV and Extended STV Geocodes have been described in enough detail to allow for constructive criticisms from a select group of persons associated with the China Historical Geographical Information Systems Project and or the Electronic Cultural Atlas Initiative. Please address any comments to: crissman@asian.gu.edu.au

L. W. Crissman
11/09/00

Appendix One

ACASIAN two letter abbreviations for contemporary Chinese provinces and districts.

Basically, this system makes use of the first pinyin letters in the first two characters of provincial-level names. However, in the cases of Heibei/Hubei, Hainan/Henan/Hunan, and Shaanxi/Shanxi, there are ambiguities, and the following additional rules are applied to create unique two letter codes that preserve the pinyin alphabetic order of the above cases: Hebei follows the normal rule, but Hubei becomes HI; if the second character/syllable is 'nan', then the second letter of the first character/syllable is used; and Shanxi follows the normal rule but Shaanxi becomes SA.

Abbreviation,	Simple Name	GB Code	GB/T 2260 – 1999 two letter code (**Bold** when different to ACASIAN)
AH	Anhui	340000	AH
BJ	Beijing	110000	BJ
CQ	Chongqing	500000	CQ
FJ	Fujian	350000	FJ
GD	Guangdong	440000	GD
GS	Gansu	620000	GS
GZ	Guizhou	520000	GZ
GX	Guangxi	450000	GX
HA	Hainan	460000	**HI**
HB	Hebei	130000	**HE**
HE	Henan	410000	**HA**
HI	Hubei	420000	**HB**
HK	Hong Kong	810000	HK
HL	Heilongjiang	230000	HL
HU	Hunan	430000	**HN**
JS	Jiangsu	320000	JS
JL	Jilin	220000	JL
JX	Jiangxi	360000	JX
LN	Liaoning	210000	LN
MC	Macao	820000*	MC*
NM	Neimenggu	150000	NM
NX	Ningxia	640000	NX
QH	Qinghai	630000	QH
SA	Shaanxi	610000	**SN**
SC	Sichuan	510000	SC
SD	Shandong	370000	SD
SH	Shanghai	310000	SH
SX	Shanxi	140000	SX
TJ	Tianjin	120000	TJ
TW	Taiwan	710000	TW
YN	Yunnan	530000	YN
XJ	Xinjiang	650000	XJ
XZ	Xizang (TAR)	540000	XZ
ZJ	Zhejiang	330000	ZJ

*Not coded or listed with an abbreviation in GB/T 2260 - 1999