CHGIS数据模型与千年尺度完整时间序列空间基础数据

——以1912年至1949年县级治所点数据为例*

路伟东

GIS(Geographic Information System)的理论研究与应用探索虽然起步很早,但直到二十世纪九十年代中期,随着PC的逐渐普及,它才真正从实验室走出来,进入到普通研究者,尤其是人文社科研究者的视野中来。在中国,因为与地理学的"亲缘"关系,历史地理学者比较早地尝试把GIS与传统历史文献结合起来,进行相关研究,①这是中国HGIS(Historical GIS)的重要源头之一。在海内外不同学科背景研究者的共同努力下,HGIS从无到有,从弱到强,不断发展壮大,成为最近二十多年来,人文社科领域最重要的学术增长点之一。②

HGIS的核心是数据,而在所有数据之中,最核心的就是"千年尺度完整时间序列空间基础数据"。2001年,国际合作项目CHGIS(Chinese Historical GIS)启动,③其最终目标,就是要构建这样的一套数据。历经十多年的艰苦研发,复旦大学CHGIS工作团队以及他们的合作者们,积累了极为丰富的基础数据,也产生了非常重要的学术影响。④但是,CHGIS的数据下限仅到1911年,这对于关注民国以来中国相关问题的研究者来讲,显然是不够的。因此,把数据的下限往后推移,尽快补全1911年之后的时空数据尤为重

* 本项目是在复旦大学历史空间综合分析实验室平台上开展的,参与项目的成员包括王新刚、陈华龙、耿金、阎芳芳、孙博、张雪峰以及朱宇飞等。本文为上海市教育发展基金会2013曙光项目(13SG10)和2015年国家社科重大项目"中国行政区划基础信息平台建设(1912—2013)"(15ZDB053)阶段成果。

① 潘威等人回顾了GIS进入中国大陆历史地理学研究的最初历程,详见潘威、孙涛、满志敏:《GIS进入历史地理学研究10年回顾》,《中国历史地理论丛》2012年第1期。

② 在这样一个过程中,台湾学者比较早地开展工作,取得了卓越的成绩,详见廖泫铭、范毅军:《中华文明时空基础架构:历史学与信息化结合的设计理念及技术应用》,《科研信息化技术与应用》2012年第4期,第17—27页。

③ CHGIS项目的详细文档,包括项目概要、版权声明、编辑机构以及数据说明、浏览、查询及下载等,看禹贡网CHGIS栏目,http://yugong.fudan.edu.cn/views/chgis_index.php?list=Y&tpid=700,2003年6月1日。

④ 根据已经发布的V4版数据统计,CHGIS共包括约65 500个点、线、面空间数据和约710余万字的释文数据。2008年中国历史地理信息系统网站(禹贡网,http://yugong.fudan.edu.cn)被美国国家人文科学基金评为人文科学优秀在线教育资源网站,http://www.sinoss.net/2008/0704/7983.html,2008年7月4日。2009年,在葛剑雄教授带领下,以CHGIS为基础研发的展示系统,曾作为教育部展厅三件实物展之一,参加了"辉煌六十年——中华人民共和国成立60周年辉煌成就展",受到教育部书面表彰,包括笔者在内的研发团队亦因此获得第八届(2011年)复旦校长奖。

要，也尤为紧迫。就目前情况而言，在可预期的时间和经费内，能够顺利推进的工作，只有县级治所点数据。本文汇报的1912—1949年数据就是这样一项工作的组成部分，希望这一数据的下限最终可以延伸至2010年，从而可以对CHGIS数据形成有效的补充，形成真正的“千年尺度完整时间序列空间基础数据”，以便更好地满足研究者的实际需要。

一、历史地理信息与千年尺度完整时间序列空间数据

任何历史事件、文化元素的发生与演变，都在特定的时间和空间中，对于研究者来讲，研究的首要前提就是要明确研究对象发生的时间和地点，也就是时间定位和空间定位。时间定位相对比较简单，中国传统历史文献纪年法，无外乎王位纪年、干支纪年或年号纪年，[①]无论哪种纪年方式，与西历公元纪年之间，都有明确的对照关系，研究者一查便知。[②]但研究对象的空间定位，就不是那么容易了。与现代地理学使用坐标投影来定义某一地理要素的确切位置不同，中国传统文献中关于地理要素空间信息的描述是通过某一特定时间切面上的点要素、面要素或线要素，即地名、辖区或河道等来进行相对定位的。

随着朝代更替，政区沿革以及河道湮废，地名辖区变化频繁，不同时间切面上的地理要素往往重叠交叉在一起，这使得那些对于记载者来讲，原本定位相当清晰的空间要素，对于后世的读者来讲，却变得模糊起来。以点(Point)要素为例，同一个地点在不同时间，甚或同一时间，往往会有不同名称。而不同时间、不同位置的点，也可能使用完全相同的名称。比如北京这个名称，现在指的是首都北京这个特定的城市，西晋时指的是洛阳，唐及五代后唐、后晋和后汉指的是太原府(今山西太原市西南)，北宋时是大名府(今河北大名县)，而金的北京则在临潢府(今内蒙古巴林左旗南)。而现在叫北京的这座城市，民国时期叫北平，历史上还曾叫过幽州、南京、燕京、大都、京师等。除了点要素，面要素(Polygon)与线要素(Polyline)的变化更为频繁，也更为复杂，从前辈学者对历史政区及河道故道的考证中，我们可以很清楚地看到这一点。[③]

对于现代的研究者来讲，这些点、线、面的中国古代基础地理信息数据，其最原始来源，大概主要是以下两个方面：

一是以《汉书·地理志》、《大清一统志》、《水经注》代表的传统文献。从《汉书·地理志》开始的16部正史地理志，每一部都用修志者所在朝代的地名重新注释前代地名。这16个时间切面上的地理信息数据，构成了相对较为完整的时间序列。除此之外，方志、总志等，也大量记载了某一地方或全国的地理信息数据。所有这些，都极大补充了正史地理志的数据缺漏。

二是以《海内华夷图》、《水经注图》、《历

① 刘乃和：《中国历史上的纪年》，海豚出版社2012年版；《文献》1983年第3、4期，1984年第1期。

② 柏杨：《中国历史年表》(上、下)，星光出版社1977年版。

③ 行政区划在GIS中是典型的面数据，河流在GIS中是典型的线数据，研究考证历史时期的政区与河道都是中国历史地理学的重要组成部分，这些问题的复杂性，从前辈学者的研究中可窥一斑。相关文章请见邹逸麟：《黄河下游河道变迁及其影响概述》《复旦学报》(社会科学版)1980年第1期；张修桂：《长江宜昌至沙市河段河床演变简史——三峡工程背景研究之一》，《复旦学报》(社会科学版)1987年第2期；周振鹤：《秦代洞庭、苍梧两郡悬想》，《复旦学报》(社会科学版)2005年第5期；满志敏：《北宋京东故道流路问题的研究》，《历史地理》第22辑，上海人民出版社2006年版，第1—9页。

史舆地图》为代表的古代地图及古代历史地图。地图是对现实世界的抽象，中国地图的绘制起源很早，也有较为完整的历史脉络。[①]不论完全写意的山水画式的地图，还是有一定规制的计里画方的地图，都有自己独特的空间表达系统和内在数学逻辑，[②]记载了大量的地理信息数据，并提供了这些空间要素相对的地理位置。

以上述文献地图为主要史料和研究对象的沿革地理学(Evolution Geography)，对历史上诸多地理现象，如各时期各政权的疆域分合伸缩，各级政区的建置沿革，各都市城邑的位置、规制与兴衰，交通路线的通塞改变，河流湖泊的变迁以及各民族的分布、盛衰和迁移等，进行了细致的复原考证。民国以来以顾颉刚、谭其骧等为代表的中国历史地理(Chinese Historical Geography)[③]学者们，又把这样的研究工作推进到一个更高的水平，不断系统化与科学化，并最终把中国传统地理信息数据，成功纳入现代地理学的科学体系框架中。

通过这些丰富、系统的传统文献以及前辈学人深入细致的研究考证，理论上，我们可以知晓某一特定地理信息数据从当代一直回溯到其最初起点的几乎所有相关信息。这些信息首先就是历史沿革，即时间序列。以县级治所点数据为例，最长的时间序列接近2 500年。时间序列外，还包括隶属关系和继承关系。隶属关系即数据层级，继承关系即数据生命开始或结束的具体情况。所有这些基础历史数据，都来源于现有的或今后的研究成果。很显然，基础数据的获取过程，实际上是就是研究的过程。

把这些传统历史地理信息数据在特定的时间精度和空间精度里考证清楚，就是千年尺度完整时间序列空间基础数据。千年尺度完整时间序列空间基础数据，就像一个沿着空间轴和时间轴展开的立体网络。在这样一个网格里，所有人文社科研究中的时空数据都可找到自己确切的位置。最终完成的这样一套千年尺度且具有完整时间序列空间基础数据，将为整个人文社科的研究，尤其是回溯的研究，提供坚实的时空参照。它实际上是了解和研究中国历史的入口和钥匙，其重要性是不言而喻的。[④]

二、CHGIS的数据模型及其理解、补充与修正

1954年冬，毛泽东批准吴晗重编改绘杨守敬《历代舆地图》的建议，成立“杨图委员会”。[⑤]1955年，工作开始后不久，谭其骧先生及其领导的工作团队就发现“重编改绘”杨图不能适应时代的要求，原设想行不通，于是及时进行了调整，最终编绘成八册《中国历史地图集》。这一调整，是把中国传统地理信息数据，纳入现代地理学科学体系框架，并真正尝试构建千年尺度完整时间序列空间基础数据工作的开始。[⑥]

《中国历史地图集》工作的思路实际上把传统地理信息数据，标准化到解放后纸质的

① 葛剑雄：《中国古代的地图测绘》，商务印书馆1998年版。

② 安敏、张春玲：《中国古代地图的数学基础与地理空间维度认知》，《测绘科学技术学报》2007年第1期。

③ 史念海：《中国历史地理学的渊源和发展》，《中国历史地理论丛》1985年第2期。

④ 葛剑雄、周筱赟：《创建世界一流应该有明确的目标——为什么要研制“中国历史地理信息系统”》，《东方学术》2002年第4期。

⑤ “杨图委员会”即“重编改绘杨守敬《历代舆地图》委员会”，葛剑雄：《悠悠长水——谭其骧前传》，华东师范大学出版社1997年版，第239—286页。

⑥ 基于近代地理坐标投影空间参考框架绘制中国地图的尝试，如《皇舆全图》、《大清一统舆图》以及民国《申报地图》等，虽然早在明末清初就已经开始，但前期所有的工作都是记载时间切面数据的地图。

测绘地图上去，这些底图主要是五十年代至六十年代的1∶200万地形图。2001年启动的中国历史地理信息系统(CHGIS)项目，[①]工作实质其实与《中国历史地图集》一样，仍然是把传统地理信息数据，标准化到当代测绘体系中来。但是，CHGIS的工作，相较于《中国历史地图集》，不论在工作的时空精度，还是数据存储方式、可视化展示等方面，都有了质的提升。CHGIS工作的底图是二十世纪九十年代百万分之一的ArcChina中国数字地图，空间精度比《中国历史地图集》提高一倍，时间精度则由每个朝代一个或几个标准时间切面[②]提高到1年。除此之外，CHGIS的最大贡献在于定义了比较合理的数据模型和数据结构，实现了传统地理信息数据在数据库中的高效存储和在GIS界面下的可视化展示。

CHGIS的数据模型来源于该项目管理委员会成员的反复讨论与集体思考，但作为这套数据直接开发者和领导者之一，复旦大学中国历史地理研究所满志敏教授在模型建立与数据最终实现方面做出了卓越的贡献。接下来对于CHGIS数据模型的讨论，均基于满志敏教授公开发表的相关文章，[③]以及CHGIS项目已经发布的第5版数据。[④]对于CHGIS模型的部分修正及补充，则来源于笔者个人参与CHGIS项目和1912—1949县级治所点数据的工作实践，以及Historical GIS的教学和应用研究。

根据CHGIS数据模型的原则性定义，数据库中最小记录单元的切分由以下三个方面主要因素决定，即：(1)地名名称；(2)行政隶属；(3)空间特征。这三个因素中的任何1个或1个以上的变化，都必须建立新的数据库记录，占据主表的一行。为了更好地描述这一数据模型的存储过程以及修正和优化，分以下三个方面进行说明：

(一) 时间序列

当代地理信息强调数据的适时性，因此，GIS数据库存储的往往是最新时间切面的截面数据，但历史地理信息最重要的特征之一是时间属性。因此，在GIS数据库中，用两维的数据表格，完整记录包含时间序列信息的三维的地理信息并不是一件容易的事情，为了解决这一问题，满志敏教授提出生存期的概念。在名称、空间特征和行政隶属保存不变的情况下，某一地理信息从出现到发生变化，通常会持续一段时间，这段时间就是地名的生存期。如图1所示。

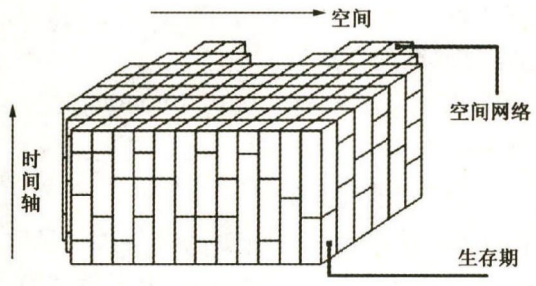

图1 时空数据概念模型

(转引自满志敏《小区域研究的信息化：数据架构及模型》)

① 葛剑雄：《中国历史地图：从传统到数字化》，《历史地理》第18辑，上海人民出版社2002年版，第1—11页。

② 谭其骧：《中国历史地图集》第1册《前言》，地图出版社1982年版。

③ 满志敏教授关于CHGIS数据模型与数据结构的讨论，主要集中在以下三篇文章中，分别是：满志敏：《地进数字化：中国历史地理信息系统的一些概念和方法》，《历史地理》第18辑，上海人民出版社2002年版，第12—22页；满志敏：《关于CHGIS第二阶段数据模型的定义问题》，《历史地理》第19辑，上海人民出版社2003年版，第231—239页；满志敏：《小区域研究的信息化：数据架构及模型》，《历史地理论丛》2008年第2期。

④ 目前CHGIS项目中文网站(复旦大学中国历史地理研究所禹贡网)公布的是CHGIS V4版数据，英文网站(Harvard CHGIS)公布的是CHGI V5版数据(http://www.fas.harvard.edu/~chgis/)。

在定义每一段生存期的开始时间和结束时间后，我们就可以把某一个特定的历史地理信息数据，切分成由前后衔接的多个生存期构成的记录集。所有的记录集共同构成数据存储的主表。而这张主表，既是所有记录集(子集)的超集，也是所有生存期记录(工作时间段，所有出现过的符合既定规则的历史地名)的超集。表1共存储了7条生存期记录，3条子集记录。对于每一个子集，给它一个统一编制的不重复的编号zjid，用于区别其他子集。并用这个编号连接一个独立编制的释文表，获取并输出属于该子集的释文。这样做，可以避免CHGIS那种为所有生存期记录编写各自的释文麻烦，不但可以减少数据冗余，也可以提供更好的用户体验，方便用户查阅该子集的所有历史沿革信息。另一方面，对于每个子集来讲，前后衔接的生存期记录，拥有不同的超集(cjid)编号。沿着时间这条主线，同一子集的生存记录，就像一串糖葫芦，可以很容易从最近的生存期记录，追溯到其生命最初的起点。

表1　主表数超集与子集

cjid	zjid	name	begyear	endyear
1	101	A	1912	1920
2	101	B	1921	1930
3	101	A	1931	1949
4	102	C	1920	1930
5	102	C	1937	1949
6	103	F	1912	1940
7	103	G	1947	1949

(二) 继承关系

实际工作中，以为简单地沿着这样一条主线，就可以追溯到所有地理信息数据生命最初的起点，是比较困难的。这是因为，绝大多数情况下，历史地理信息的沿革变化都是树形的，而非单一轴线式的。它们可能产生于一个，或多个树根之中，在生长的过程中，不断的分叉，形成子树的同时，也有子树生命的终结。如图2所示，树1单一轴线型的A—B—C—D在主表中，是由4条连续的生存期记录构成的子集，拥有共同的子集(zjid)编号。而E在从A分出后，因为没有取得和B一样的正统地位，它在主表中是一条全新的记录。E和E的子节点F使用另外的子集(zjid)编号。在这种情况下，从F开始的回溯，在数据表里只能到达E，而无法反映父节点A的信息；G记录也很有代表性，它来自E，但同样没有正统地位，在主表中是一条新的记录，拥有自己的子集(zjid)编号。另一方面，G虽然有两个子节点H和I，但这两个子节点没有一个继承G的正统，而是瓜分了G。因此，在主表中，H和I又都是全新的记录，都拥有各自独立的子集(zjid)。而G的生命就终结于它自己，在主表中只有一条生存期记录。于是从G、H、I开始的回溯，都只能到达它们自己，而无法反映它们各自父节点的信息。

树2的情况也很有意思，来源于三个父节点的D是主表中全新的记录，和它的子节点E组成一个子集。但从E开始的回溯，仍然只能到达D，而无法反映D的父节点信息。

满志敏教授的文章中没有讨论这方面的内容，已经公开发布的CHGIS数据也没有相关的表单设计。尽管如此，CHGIS数据模型和数据结构设计之初，应该考虑了这种情况，只不过可能因为各种原因，没有体现在最终的数据中。其实解决这一问题非常简单，只要建立一个单独的继承关系表就可以了。

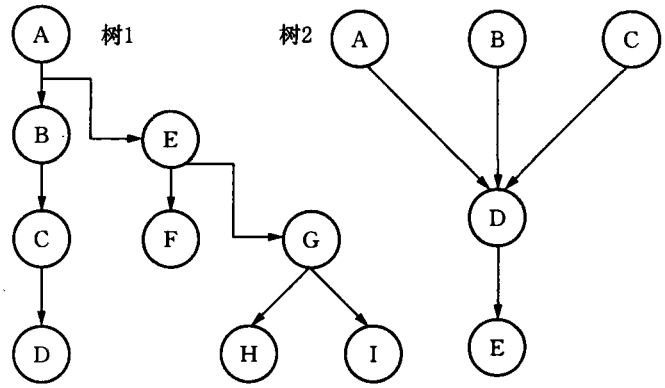

图 2　历史地理信息数据沿革树

如表 2 所示，对于需要建立继承关系的记录，我们只要继承关系表中标注这条记录在主表中的超集（cjid）编号，和其父节点的超集（cjid）编号即可。对于父节点有多个来源的，如图 2 树 2 中的 D，只要在继承关系表中添加三条父节点记录即可。在回溯主表中某一子集的数据时，当指针到达时间最早的一个生存期记录后，再以 cjid 连接并查询继承关系表，如果有查询记录，就把该记录，添加在该子集记录的前面，一并输出。

表 2　继承关系表字段说明

字段名	字　段　说　明
id	数据库自动编号
cjid	主表中需要添加继承关系记录的超集 ID
jcid	主表中需要添加继承关系记录的父节点超集 ID
isborn	布尔值，用于标注该记录是记录开始还是终结
year	继承关系开始的年份

某县设置之后，在中间的某一时段内，可能曾经撤销过，后又复置，这种情况是存在的。体现在主表中，相应子集的多条生存期在时间上就有可能是不连续的。在这种情况下，如果进行回溯，我们可能就无法判断，继承关系表中查询到记录，应该添加在主表子集的具体位置。因此，需要在继承关系表中增加时间字段，用于标识。

实际上，继承关系表还可以存储主表生存期记录生命终结的信息，比如某一县被裁撤之后，是否被合并于其他县中，或被分割到其他多个县中。记录的方式和前面一样，只不过我们需要在继承关系表中再添加一个布尔值的字段，用以标识该条记录的属性。

（三）隶属关系

隶属关系描述的是某一历史地理信息数据的上属，当这一隶属关系处于一个时间序列中时，其上属实际上是一个数据集。比如某一县，历史上可能曾经隶属于不同的州、路、府、厅等。CHGIS 数据模型原则定义行政隶属关系变化应该在主表中添加新的记录，但行政隶属的变化，实际上往往和数据本身没有关系。满志敏教授的文章中列举了一个比较极端的案例，於𥫱县。该县始置于公元前 221 年，公元 25 年改为於潜县，此后直至 1911 年共 1 886 年间，其名称和治所位置均没有发生变化，但其行政隶属在这期间，却变化了 21 次。很显然，在主表中，用 2 条生存期记录比用仅仅为了体现其上属变化而切分成 22 条的生存期记录更简洁，更合理，也更有效。

那如何解决这一问题呢？CHGIS给出了两种可能的方案，即：(1)利用GIS查询功能，检索空间包含关系，再给出行政隶属关系。(2)建立单独的关系表，列举地名所有的行政隶属关系。第一种方法实际上是GIS基于拓扑关系的空间查询，实践表明，这种查询的缺点不只是没有给出直接的行政隶属关系表，而是效率极低，尤其当面临较大数据量时，几乎是不可接受的。实际工作中，CHGIS没有采用这一方案，而是采用了第二种方案，即通过建立主表生存期记录与行政上级一对一关联的串连表，来记录隶属关系。实际上，当行政上属的时间跨越一个以上的生存期记录时，仍然要根据各生存期记录的起讫时间，来切分行政上属的记录。如图3所示，这种一对一的数据存储方式，虽然可以精确还原主表生存期记录的隶属关系，但却增加隶属关系表的重复记录，产生了冗余数据。更麻烦的是，当主表生存期记录发生变化时，必须在隶属关系中做出相应的修改，以保持数据的一致性，这显然会带来修改的不便，且容易产生错误数据。

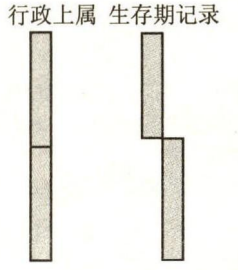

图3 生存期记录与上属

表3 隶属关系表字段说明

字段名	字 段 说 明
id	数据库自动编号
zjid	主表中需要添加隶属关系记录的子集ID
ssid	主表中对应子集的行政上属ID
begyear	该行政上属开始的年份
endyear	该行政上属结束的年份

实际上，隶属关系表中的行政上属，完全没有必要根据生存期记录时间进行切分。优化方案其实非常简单，只要我们用主表的子集(zjid)编号，而不是用超集(cjid)编号来定义隶属关系表，就可以解决这一问题。如表3所示，在隶属关系表中，我们可以用生存期的定义，记录每一个子集所有的行政上属及其出生和死亡的时间。在查询主表各生存期记录的上属时，我们只需用SQL语句直接把符合时间要求的行政上属分配到各生存期记录中去即可。

三、1912—1949年数据说明与数据展示

1912—1949年县级治所点数据是“1992年至2013县级治所时间序列点(TSP，Time Series Points)数据”这一长期计划的第一阶段数据，时间上与CHGIS县级治所TS数据(只覆盖到1911年)相衔接。该工作最终完成后，有关中国县级治所的时间序列将可以从21世纪初，一直回溯到公元前5世纪，前后长达近2 500年。因此，这一数据，是构建“千年尺度完整时间序列空间基础数据”的重要组成部分。

这一工作得以开展，首先基于民国政区已有的研究成果，尤其是傅林祥、[①]郑宝恒[②]两位先生的研究。系统汇总梳理这些已有的研究，我们基本可以理清了民国时期县级政区的历史沿革，这为1912—1949年县级治所点数据开发提供了最基本，也最重要的地理信息数据支持。

① 傅林祥、郑宝恒：《中国行政区划通史·中华民国卷》，复旦大学出版社2007年版。

② 郑宝恒：《民国时期政区沿革》，湖北教育出版社2000年版。

具体工作中，除了上节所述对 CHGIS 据模型和数据库设计等方面有部分的修正与补充外，还有部分重要指标与 CHGIS 存在差异，需在此说明：

（一）关于工作底图、坐标与投影

CHGIS 的工作底图是 ESRI 公司发布的 Digital Map Database of China，该图实际由中国国家测绘管理局制作，又称 ArcChina。根据 CHGIS 网站公开发布的信息，ArcChina 原始数据大地基准面（Datum）是 Pulkovo_1942，参考椭球体（Spheroid）Krasovsky_1940，精度百万分之一。①实际上，自二十世纪五十年代初开始，中国以 Pulkovo_1942 为基础进行了联测，建立了自己的北京 54 大地坐标系。②所以，由中国官方制作并发布的 ArcChina 大地基准面应该是北京 54，而非是 Pulkovo_1942。CHGIS V4 版提供了西安 80 的 ArcChina 底图数据，这显然是在北京 54 基础上转换而来。

北京 54 坐标系虽然根据我国实际情况进行了平差计算，但实际上只不过是 Pulkovo_1942 的延伸，其测量起始的大地原点和数据表达的核心区域，不论经度还是纬度都与中国相去甚远，使用这种大地基准面来记录中国的数据，本身存在较多的问题。西安 80 虽然定义了中国自己的大地原点，并且采用 IUG 1975 椭球体，但却与北京 54 一样，没有公开转换参数。这使得这两种坐标系与实际的国际标准 WGS84 之间没有现成的公式可以完成转换。③更重要的是，ArcChina 数据仅覆盖中国现代边界以内区域，而千年尺度完整时间序列空间数据所覆盖的区域是历史时期的中国全境，④其范围要远大于现代中国。因此，使用 ArcChina 作为工作底图，意味着有相当一部分地区将缺少标准数据。而且，ESRI 目前已经停止分发新的 ArcChina 数据包了。很显然，对于构建面向全球用户的千年尺度完整时间序列空间基础数据来讲，不论使用北京 54 还是西安 80 的 ArcChina 底图，都存在较多问题，不是最优选择。

与众多人为定义的大地原点相比，客观存在的地球质心只有一个。所以，以地心为坐标原点的 WGS84 坐标系，不存在歧义，且参数公开，与世界上绝大多数的大地坐标之间，没有数据转换问题，是实际上的全球标准。目前使用最广泛的全球定位系统 GPS 和最知名的虚拟地球 GE（Google Earth）均采用 WGS84 坐标系，后者也是一般研究者可以免费获取高精度全球影像数据的最重要来源之一。GE 球状卫星图像整合界面极其简单友好，为用户提供了前所未有的方便。虽然各地区精度不一，但信息完整，且总体细节够多。更重要的是，GE 提供的地形、海拔、经纬度信息和手持 GPS 输出的经纬度信息，是一致的，这对于以实体考察为重要手段

① Harvard CHGIS, http://www.fas.harvard.edu/~chgis/data/arc_china.htm.

② 吕志平、许东周、朱华统、沈明顺：《我国高精度地心坐标转换参数的建立》，《解放军测绘学院学报》1995 年第 2 期。

③ 王解先、王军、陆彩萍：《WGS-84 与北京 54 坐标的转换问题》，《大地测量与地球动力学》2003 年第 3 期。

④ 谭其骧对历史时期中国的范围进行了系统的表述，他指出“十八世纪五十年代清朝完成统一之后，十九世纪四十年代帝国主义入侵以前的中国版图，是几千年来历史发展所形成的中国的范围。历史时期所有在这个范围之内活动的民族，都是中国史上的民族，他们所建立的政权，都是历史上中国的一部分。”（谭其骧：《〈中国历史地图集〉总编例》，谭其骧主编：《中国历史地图集》第 1 册《原始社会·夏·商·西周·春秋·战国时期》，地图出版社 1982 年版。）千年尺度完整时间序列空间数据的覆盖范围就是谭其骧定义的历史中国的范围。

的历史空间数据的考证来说，具有很高的实用价值。[①]与传统历史地理考证相比，传统文献与 GE 影像数据的结合研究，可以极大提高历史空间数据定位的准确性，也能极大提高工作的效率。

基于此种考量，本项目不采用北京 54 或西安 80 的 ArcChina 工作底图，也不提供 PCS(Projected Coordinate System)投影之后的平面坐标数据，而直接使用 GE 读取历史空间数据的经纬度数值，最终仅提供 GCS 投影的 WGS84 球面坐标数据。世界各地用户可以根据各自不同情况，自行转换满足实际需要的 PCS 投影数据。

（二）关于工作空间精度和时间精度

CHGIS 数据的空间精度是 1∶100 万，对于建设一个以县级治所为最小点单位的历史地理信息数据库来讲，这一空间精度是能够满足要求的。但是，越靠近当代，数据容量越大，数据层级越多，要求的精度也越高。而研究者自己的数据，比如道路、渡口、关津、寺庙、学校、井泉等等，在数据容量和精度等方面同样如此。因此，目前所展开的县级治所点数据只是其中的一个数据图层，民国以来完整的历史空间基础数据还应当包括县以下聚落及其他相关点、线、面数据。

地图比例尺的大小决定着实地范围在地图上缩小的程度，比如 1 平方公里面积的聚落点，在百万分之一地图上为 1 平方毫米，仅能表示成一个点。但在十万分之一地图上为 1 平方厘米，可以表示出该居民地的大体轮廓以及与最近河流、道路等线要素的相对位置。同时，在表达更低层级聚落点及相关信息时，也留有足够的存储空间。考虑到数据今后的可扩展性，将 1912—1949 年县级治所点数据的工作精度在 CHGIS 的基础上提高一个数量级，定为 1∶10 万，是比较合适的。

GE 比例尺在视图菜单里可以打开，窗口显示的是 60 度经典视角的相机图像，右下角的眼球海拔(Eye Alt)高程实际就是该相机高程，该标高减去图像所在标高才是相对地面的视点高度。所以，GE 窗口平移图像时，海拔数据会发生变化，比例尺亦会变化。使用 Alt 键加上数字键盘的＋或者－键，可以以 1 米为单位微调至所需要的比例尺。但实际工作中，为提高工作效率，直接由鼠标滚轮调整比例尺，精度误差控制在正负 50 米以内。

本项目工作的时间精度为年，不同类型时间标签的处理方式沿用 CHGIS 的成例，在此不多做说明。

（三）关于规则说明

因为各种原因，在中国大陆地区，GE 影像与地名标注存在系统性误差，在部分地区这种误差还相当大。由于这种误差是人为刻意造成的，进行有效的校正难度较大。因此，本项目空间数据的定位统一以 GE 影像数据为准，且规定行政治所为最终取值位置。

本项目还定义了空间数据的开始规则与结束规则，包括数据上限、新建、更名、迁移治所、更名并迁移治所、裁撤、合并、复置、数据下限等等，基本沿用 CHGIS 的成例。所需说明者，在史料判读上，时间序列的切分和新记录的添加，原则上以官方公报为准，凡未经官方正式书面批准的变更行为，一般均不在数据库中添加新的记录，而仅在释文中加以详述。比如甘肃宁定县，民国 6 年(1917 年)七月析导河县地置，于同年 10 月成立县署。

① 林选妙、黄丽蓉、张兴、陈雅芳：《Google Earth 在全国地名普查项目中的应用》，《大众科技》2013 年第 1 期。

因甘肃省政府未报县署成立日期，内务部未向大总统转呈。直到民国 8 年(1919 年)三月，内务部、财政部始会呈照准。[①]因此，切分宁定县时间序列应该在 1919 年，而非 1917 年。

官方批准后未实际执行，不增加新的记录。比如宁夏磴口县，民国 31 年(1942 年)三月，国民政府核准迁治广兴源，[②]但直至 1949 年解放前，县政府仍在磴口，[③]即今之巴彦木仁苏木，并未真正迁移。因此，1942 年之后的磴口县在数据库中不会切分为新的记录；广西资源县情况类似，虽然民国 13 年(1924 年)3 月，广西省议会就议决析置。[④]同年 10 月，亦已经内务、财政两部呈准。[⑤]但实际并未实行，至民国 25 年(1936 年)7 月，始重新划定区域，始析全县西延区、长万区以及兴安县北部车田、寻源 2 乡置。因此，在数据表中，资源县第一条记录始于 1936 年。

浙江文成县的情况更复杂，民国 35 年(1936 年)12 月核准析瑞安、泰顺、青田三县交界地置，但奉准暂缓成立。[⑥]直至民国 37 年(1948 年)7 月始正式成立。时定县政府驻黄坦，在大坐镇设办事处，但因大坐交通较黄坦方便，县政府实际设在大坐。[⑦]规则定义，凡实际位置于批准位置不一致者，以实际位置为准。因此，在数据表中，文成县第一条记录始于 1948 年，而其治所则定位于大坐。

1912—1949 年间，时局多变，政区亦受其影响，其过程相当复杂，如日伪占据，解放区国统区拉锯，边疆地区分离，等等。对于个别时间序列切分有异议的县，视具体情况而定，坚持上述原则，可以避免很多不必要的麻烦。如宛平县民国 19 年(1930 年)初迁卢沟桥拱极城内清代西路厅旧署(今北京市西南郊卢沟桥)，日伪时期迁驻长辛店，抗战胜利后因之。[⑧]又如河南武陟县，初治今河南武陟县驻地木城镇西南、阳城乡东北沁河河道中间。民国 28 年(1939 年)，日军入侵武陟，在木栾店(今武陟县驻地木城镇)置伪武陟县公署，抗战胜利后，流亡乡间的武陟县政府改驻木栾店。[⑨]两县时间序列切分皆应在抗战胜利后，即 1945 年。

即便如此，实际工作中，仍有很多棘手问题。比如地名读音、地名字形、地名字体、地名类别以及特殊县级政区等等，都需要认真对待，并在数据库表中加以体现和说明。除此之外，作为 1912 年至 2013 年长时段数据的第一期工作，该数据实际上具有一定的实验性质，既与 CHGIS 原有的工作实践有一定的传承，也有部分自己的改进。与此同时，在继续开展进行下一步的工作之前，可能还有较多问题，尤其是在数据的标准与规范方面，仍然需要深入思考和不断调整。这些问题包括，但不限于，地理编码、数据平台、坐标系统以及基础底图，等等。[⑩]

---

① 《政府公报》第 1120 号，1919 年 3 月 18 日，第 83 册，第 548 页。

② 《政府公报》渝字第 448 号，1942 年 3 月 14 日，第 15 页。

③ 赵钟贤：《回忆共产党接管国民磴口县政府前的历史背景及接管过程》，《磴口文史资料》第 11 辑，1994 年印，第 1 页。

④ 吴承堤：《近六十年全国郡县增建志要》卷下，第 19 页。

⑤ 《政府公报》第 3074 号，1924 年 10 月 14 日，第 150 册，第 4399 页。

⑥ 内政部方域司：《中华民国行政区域简表》(第 11 版)，第 19 页。

⑦ 《文成县志》，中华书局 1996 年版，第 15、984 页。

⑧ 《北京市丰台区志》，北京出版社 2001 年版，第 187 页。

⑨ 《武陟县志》，中州古籍出版社 1993 年版，第 51 页。

⑩ 在 2015 年 4 月份召开的第一次历史地理信息系统(HGIS)学术沙龙和 2015 年 12 月召开的国家社科基金重大项目“中国行政区划基础信息平台建设(1912—2013)”(15ZDB053)开题论证会上，广州大学地理科学学院吴志锋教授在数据规范和标准方面给予较多指导和帮忙，在此致谢。

四、余　论

任何历史人物、事件与文化要素都在特定的时间和空间中发生、演变，都是人文社会科学工作者的研究对象。用 GIS 的术语来表述，它们都是具体时间特征的空间数据。对这些历史人物、事件和文化元素进行准确、快速的时间和空间定位，是继续研究的首要前提。而千年尺度完整时间序列空间基础数据，像一张沿着时间轴，往历史方向延伸的立体的大网，为研究提供了极其重要的背景空间数据，为这样的时空定位工作提供了基础参照系。

对于研究者来讲，千年尺度完整时间序列空间数据中精确的面状数据固然重要，但从工作的难易程度上看，点数据更具有可实现性，而足够多的具有完整时间序列的点数据，实际上，已经可以满足大部分的研究需要。1912—1949 年县级治所点数据的工作实践表明，在可预期的时间和经费内，构建千年尺度完整时间序列点数据，尤其是县级治所点数据，是可行的。

HGIS 的核心是数据，但 HGIS 的灵魂是空间分析，是要把自己的研究对象，放置在历史时空间的框架中，重新去审视它们，研究它们。以期发现，并解决那些仅仅通过传统文字描述或简单数学统计，无法发现和解决的问题。从这一角度讲，以千年尺度完整时间序列空间数据为基础核心数据的 HGIS，是大数据时代人文社科数据整合最佳平台，更是综合交叉研究的最佳平台。