

数据库建设与清史研究

胡 恒

[摘要] 清史数据库建设在近二十年内取得了突出的成绩，科研机构的推动与商业公司的运营是两大动力。清史数据库建设及随之而来的量化史学研究克服了传统研究缺乏系统性数据或数据缺失严重等弊端，为计量方法的应用提供了相当便利的条件，但也将带来史料运用与理解方面一系列新的问题。为应对清史数字化发展进程，有必要建设数据共享平台，并将公益与商业开发适度结合，但与此同时，在数字时代更要重视传统史学的训练。

[关键词] 资料库 数字人文 清史研究

[中图分类号] K249 **[文献标识码]** A **[文章编号]** 1002-8587(2016)-04-0017-09

自20世纪90年代以来，随着信息化、电子化、网络化给人类社会带来的全面冲击，史学领域数据库建设开始逐渐兴起，一批重要的研究史料陆续以各种形式电子化并制作成服务于不同对象的数据库，从而极大地改变了史学的生态，引起了史学信息获取方式的变革。如今多数年轻学人已习惯于通过数据库的形式获取学术资源，从而使得年轻一代的史学写作与前代学者相比，出现了一定程度的代际差异。对这一变革带来的利弊，学术界仍然存在相当大的分歧。^①

清朝是距今最近的一个传统王朝，向来以史料丰富著称。自21世纪初启动“清史工程”以后，清史文献的整理成为四大基础工程之一，一大批文献资料被制作成数据库或影印出版，促进了整个清史学界在21世纪初十多年的持续繁荣，目前已建与在建的关于清史资料的数据库在各个断代史中数量可能是最多的。对清史研究而言，数据库以及随之而来的量化研究因在一定程度上克服了之前各断代常常存在的缺乏系统性数据、数据缺失严重等弊端，具备一定的“大数据”基础，为数据库的建设与计量方法的应用提供了相当优越的条件。这一领域的研究虽进展迅速，但迄今为止，尚未得到系统的总结。^②本文仅就笔者目力所及，针对数据库建设与清史研究的关系谈一点浅见，难免挂一漏万，不当之处，敬祈方家指正！需要说明的是，本文主要针对的是目

[收稿日期] 2016-05-20

[作者简介] 胡恒（1984-），男，中国人民大学清史研究所副教授，北京100872；huhengqss@ruc.edu.cn

[基金项目] 国家社科基金青年项目（编号：16C25018）阶段性成果。

① 见梁晨、董浩、李中清：《量化数据库与历史研究》，《历史研究》2015年第2期。有关古籍数字化及数据库方面的理论与实践，可参考毛建军《古籍数字化理论与实践》一书，航空工业出版社，2009年。关于历史研究数据库方法运用的理论阐释，可参金观涛、刘青峰：《历史的真实性：试论数据库新方法在历史研究中的应用》，《清史研究》2008年第1期；项洁主编：《数位人文在历史学研究中的应用》，台大出版中心，2011年。

② 关于灾害史方面，夏明方教授作了很全面的回顾与总结，可参《大数据与生态史：中国灾害史料整理与数据库建设》，《清史研究》2015年第2期。

前已上线的较重要的各类数据库，未公开发布的不在之列。

—

清史研究的最大优势在于史料的丰富。戴逸教授为“清史工程”《文献丛刊》所做的“序言”中，曾将清史文献的特点概括为“多、乱、散、新”，文献典籍“至清代而浩如烟海，难窥其涯矣”^①。仅就刊印下来的典籍而言，已堪称汗牛充栋，更不必提及大量以手稿、文书形式存在的未刊文献。

为便于检索、利用清史资料，早期学术界往往通过编写索引、年表等工具书的形式来方便搜集信息。这类工具书分为两类，一类是检索型工具书，一般只提供查找文献信息的线索，不直接回答所要查询的内容，主要指书目、索引等；一类是参考型工具书，一般不提供资料线索，仅简要提供回答问题的知识，主要指各类字典、词典、类书、年鉴、图录、表谱等等。^②当时学术界比较重视工具书的编纂，也取得了很大的成绩，比较重要的检索型工具书有中国社会科学院历史研究所与中国人民大学清史研究所合编的《清史论文索引》（中华书局，1984年）、杨廷福、杨同甫编《清人室名别称字号索引》（上海古籍出版社，1988年）、中国人民大学清史研究所、中国社会科学院边疆史地研究中心合编的《清代边疆史地论著索引》（中国人民大学出版社，1988年）等等；参考型工具书有章伯锋编《清代各地将军都统大臣等年表（1796—1911）》（中华书局，1965年）、魏秀梅编《清季职官表（附人物录）》（中研院近代史研究所，1977年）、钱实甫编《清代职官年表》（中华书局，1980年）、《中国历史大辞典》（清史卷）（上海辞书出版社，1992年）、朱金甫、张书才主编的《清代典章制度辞典》（中国人民大学出版社，2011年）、中国人民大学清史研究所资料室编《清代中西历表（1573—1840）》（中国人民大学出版社，1980年）、江庆柏编《清代人物生卒年表》（人民文学出版社，2005年）等。^③这些工具书，以及用以记录资料的卡片，与今天的数据库具有很大的相似性，只是表现形式差别较大而已。那时也开始利用这年年表做一些量化分析，只不过数理统计依赖于手工，如魏秀梅利用《清季职官表》，连续在中研院《近代史研究所集刊》上发表了《从量的观察探讨清季布政使的人事嬗递现象》（第二期，1971年）、《从量的观察探讨清季按察使的人事嬗递现象》（第三期下册，1972年）、《从量的观察探讨清季督抚的人事嬗递》（第四期上册，1973年）等文章。

20世纪80年代以后，随着信息技术的发展，以网络环境搭建清史数据库的工作就开始了，台湾是这一领域的先行者。1984年台湾中研院历史语言研究所为“史籍自动化”计划的延伸而开始了“汉籍全文资料库计划”，目标是收录对中国传统人文研究具有重要价值的文献^④，其中，清史文献成为这一计划的重要组成部分。这一领域标志性的事件是1999年上海世纪出版集团、上海人民出版社、香港迪志文化出版有限公司推出的《文渊阁四库全书》电子版，迅速受到学术界的欢迎。

21世纪以后，随着网络技术的发展，尤其是个人电脑用户的剧增，使得数据库的建设迎来了前所未有的发展契机，史学类的数据库建设也日渐兴盛。清史研究资料的庞大，为数据库的大规模建设提供了良好的契机。借助文献资料的大规模整理与检索手段的引入，学者可以更快地掌

① 戴逸：《清史〈文献丛刊〉、〈档案丛刊〉总序》，载朱诚如、王天有主编：《明清论丛》第5辑，紫禁城出版社，2004年。

② 彭莲好、王勇主编，朱宁副主编：《现代信息检索基础教程》，华中科技大学出版社，2014年，第38页。

③ 黄爱平主编的《清史书目》（1911—2011）（中国人民大学出版社，2014年）专门列有“工具书”一项，系统收录了1911年至2011年百年间与清史相关的综述、地图、年表等工具书的书目。

④ 见“汉籍电子文献资料库”的说明，<http://hanchi.ihp.sinica.edu.tw/ihp/hanji.htm>，访问时间：2016-09-01。

握某一领域较为全面、丰富的学术资源。清史数据库建设还有另外一个优势在于系统性史料的留存较为显著，具有较好的数据基础。这些数据相对来说制作标准比较统一，将其处理为数据集较为容易。譬如清代户部所积累的人口、土地数字就是比较连续的，粮价数据可为应用现代经济模型提供基础，雨雪分寸的资料可以为研究清代气候变化提供相当精准的定量基础，《缙绅录》提供了相当完整的清代官员的出身、籍贯、任职时间等标准化的数据。

因是之故，近一二十年来，清史数据库建设发展迅速。台湾起步较早，大陆近些年发展迅速，海外数据资源零散但十分丰富。通常对数据库分为检索与计量两类^①，目前检索型数据库数量最多，已形成了基本文献数据库与专题文献数据库齐头并进，共同发展的繁荣态势。所谓的基本文献库，指的是清史领域最基本、最常用、最重要的文献及其集成，包括档案、实录、政书、地方史志等等；专题文献数据库则多以纸本资料集为依托并结合学术界已有的研究而制作，如清代权威人名档等等，其中既有若干档案保管机构出于档案整理的便利而进行的数据库建设，也包括各科研单位基于自身研究优势所建设的各类专题数据库。前者覆盖面广，几乎为多数清史研究工作者所使用；后者则专业与特色比较明显，往往建立在前期丰硕的研究成果之上。

二

档案是清史研究的生命。台北中研院史语所的“明清档案工作室”从1996年开始将所藏“内阁大库档案”数字化，2001年起提供网络线上阅览和复印服务^②，只是目前大陆地区开通账号使用权限的高校与科研院所还为数甚少，但可以进行档案目录的检索。台北故宫保存的档案资源十分丰富，该机构早年即已开始数字化工作，并建立了相当完备、自成体系的数据库系统，包括清代宫中档奏折及军机处档折件全文影像资料库、清宫奏折档台湾史料目录、清代权威人名档案资料查询、大清国史人物列传及史馆档传包传稿资料库、清代文献档案总目、同人辑清史书目。台北故宫的数据库是以该馆丰富的清史档案馆藏为基础的，收录的多数资料为该馆所独有，且设计精良、界面友好、权威性强，目前已成为清史工作者必须常常查阅的权威数据库。该数据库在大陆高校和科研单位使用率尚不高，多数无法进行全文影像的查阅和复印，但可以进行档案目录的检索。此外，台北故宫还对中国第一历史档案馆所藏的清代满汉文朱批奏折汇编、活计档作名建立了目录检索系统。^③

中国第一历史档案馆是清宫档案最大的收藏机构，从本世纪初清史工程启动以后，中国第一历史档案馆将为纂修清史提供馆存档案整理作为工作的重中之重，为此专门设立整理规划项目，对该馆所藏档案进行系统的整理、编目、扫描、拍照工作。从2006年10月开始为期一年的清代档案数字化试点工作，又成立了“清代档案数字化试点工作小组”等机构，制定了《中国第一历史档案馆清代档案数字化试点工作实施方案》^④。经过近十年的努力，中国第一历史档案馆的数字化工作取得的进展是非常明显的，阅档条件也在逐步改善。目前，该馆馆内已建立了可供查阅的档案系统，其中收录的数据库主要有该馆所藏的朱批奏折和录副奏折，也包括部分户科题

① 项洁、翁稷安：《数位人文和历史研究》（载《数位人文在历史学研究的应用》，台大出版中心，2011年），所谓检索，指的是数位典藏和资料库的发展，计量指的是计量史学所进行的统计实验，第13页。

② 见明清档案工作室官网介绍，<http://archive.ihp.sinica.edu.tw/>，访问时间：2016-09-01。另可参洪一梅：《明清内阁大库档案数位管理机制：整合式档案管理自动化系统》，2005年6月《第三届两岸三院信息技术与应用交流研讨会论文集》。

③ 台北故宫博物院“典藏资料库系统”网址是：<http://www.npm.gov.tw/zh-TW/Article.aspx?sNo=02000021>，访问时间：2016-09-01。

④ 中国第一历史档案馆：《清代档案数字化试点工作总结报告》，载《档案信息资源开发利用试点经验汇编》，中国档案出版社，2008年。

本、刑科题本、上谕档、随手档，也包括清代灾赈档、内务府奏案等，《清实录》与《清会典》等史料也同机提供检索^①，为研究者提供了较大的便利。目前这批档案同时在国家清史编纂委员会“数字图书馆”中收录。只是两者利用起来仍有很多不便，前者只能在中国第一历史档案馆馆内查阅，无法远程登录，对于京外学者尤其不便；后者则主要是为参加清史纂修工作的专家提供阅档服务，涵盖范围极其有限。将来条件成熟时，如能将这一系统建设为可远程登录、查询的开放型数据库，相信对整个清史研究的推动意义是无法估量的。值得一提的是，中国第一历史档案馆从2014年起逐步在官方网站上分期分批开放馆藏数字化档案目录，首批在网站已推出汉文朱批奏折档案目录^②，尽管离学术界的需求还有很大距离，但毕竟已是一个可喜的进步。另据爱如生公司官网介绍，他们将推出“明清档案库”，该库汇辑海峡两岸陆续公布的明清官私档案，包括宫中档、理藩院档等，也包括各部衙公务文书、各州府赋役清册、民间档案等，拟分为5集，2017年起陆续出版^③。

与此同时，地方档案的数字化工作取得了显著进步。清代的县级档案保存下来的比较重要的有淡新档案、巴县档案、南部县档案、获鹿县档案、顺天府档案等^④。其中淡新档案保存在台湾大学图书馆^⑤，边整理边出版边数字化，被纳入项洁教授主持的台湾历史数位图书馆（THDL）内，形成比较完备的淡新档案的数据库，且对公众开放^⑥，这是一项堪称典范的数据库建设的样本。巴县档案早已整理，出版过《清代乾嘉道巴县档案选编》^⑦、《清代巴县档案汇编（乾隆卷）》（档案出版社，1991年）、《清代四川巴县衙门咸丰朝档案选编》（上海古籍出版社，2011年）、《清代巴县档案整理初编 司法卷（乾隆朝）》（西南交通大学出版社，2015年）等，在此基础上，建立了巴县档案的数据库，四川省档案局在网上提供巴县档案目录检索^⑧。四川南部县所保存的清代档案极其丰富，从顺治十三年直至宣统三年兵房、吏房、工房、户房、刑房、礼房、盐房的档案保存都比较完整，对研究清代地方司法、行政等均有极大的史料价值，备受海内外学界关注。南部档案的数字化整理与出版近年被纳入资助项目，先后出版了《清代南部县衙档案目录》（中华书局，2009年）、《清代四川南部县衙门档案》（黄山书社，2016年）。巴县档案与南部县档案在中华文史网“数字图书馆”中可进行档案查询和浏览，只是开放程度还相当有限。顺天府档案在中国第一历史档案馆提供胶片阅览。此外，阿拉善左旗档案馆所藏阿拉善档案、青海省档案馆所藏循化厅档案在中华文史网的“数字图书馆”也提供目录检索与全文浏览，但存在的主要问题是开放程度有限。

① 《中国第一历史档案馆开放浏览胶片、数字化档案目录》，<http://www.lsdag.cn/cdxz.jhtml>，访问时间：2015-08-01。

② <http://www.lsdag.cn/dacx/index.jhtml>，访问时间：2016-09-01。

③ 爱如生官网，http://er07.com/home/pro_15.html，访问时间：2016-09-01。

④ 参见胡忠良：《全国各地档案馆所藏清代档案基本情况调查报告》，载中国第一历史档案馆官网，<http://www.lsdag.com/yjbg/709.jhtml>，访问时间：2015-08-01。该报告为清史纂修工程启动以后，清史编纂委员会为对全国清代档案保存情况进行的摸底调查，报告共五份，未刊，包括：《中国第一历史档案馆所藏清代档案基本情况调查报告》、《散失在境外清代档案文献调查报告》、《清代画图收藏情况和价值调查报告》、《全国清代档案出版情况调查报告》，见马大正：《开局顺利 任重道远——清史纂修工程简介》，中国史学会秘书处、陕西师范大学历史文化学院编《中国历史学研究现状和发展趋势》，中国社会科学出版社，2006年。

⑤ 可参《台湾“淡新档案”与所谓台湾“旧惯”之搜集与整理》中的《淡新档案介绍》，载黄静嘉《中国法制史论述丛稿》，清华大学出版社，2006年。

⑥ <http://thdl.ntu.edu.tw/index.html>，访问时间：2016-09-01。

⑦ 由四川省档案馆、四川大学历史系主编，上册于1989年、下册于1996年由四川大学出版社出版。

⑧ http://221.10.28.84:22343/hxtext_form_scda/website/scda/arch!getInfoByArchName.action，访问时间：2015-08-01。

《清实录》与《清会典》是清史研究的基本史料，与档案相比，其更加系统与全面，不少档案需要借助实录相关条文的记载才得以明确其上奏的背景，因此，即使清代档案保存的数量如此之多，《清实录》与《清会典》仍然具有非常重要的史料价值。在清代档案利用尚未成熟的过往清史研究历史上，《清实录》与《清会典》是最常用的史料集，尤其是《清实录》，由于部头庞大，不便利用，以往学界根据需要，分门别类制作了各类专题摘编，如《〈清实录〉北京史料辑要》（紫禁城出版社，1990年）、《清实录藏族史料》（西藏人民出版社，1982年）、《〈清实录〉准噶尔史料摘编》（新疆人民出版社，1986年）、《清实录经济史资料》（北京大学出版社，1989年）、《清实录中朝关系史料摘编》（吉林文史出版社，1991年）、《清实录类纂·科学技术卷》（武汉出版社，2005年）等等，至少有十余种之多。目前这样分类摘编的工作已意义不大，中国第一历史档案馆与北京书同文数字化技术有限公司合作制作了“大清历朝实录”和“大清五朝会典”数字化成果，实现了全文可检索，中研院与韩国国史编纂委员会合作开发的“明实录、朝鲜王朝实录、清实录”资料库最近也已免费开放^①。

大型资料集中有关清史的内容也非常丰富。比较重要的除了《文渊阁四库全书》电子版外，还有台湾中研院的“汉籍电子文献资料库”（又称“瀚典”全文检索系统，与清代有关的子系统有汉籍全文资料库、台湾文献丛刊、清代经世文编系列、内阁汉文题本专题档案：刑科婚姻类提要、新清史—本纪等）。爱如生公司建立的十多个古籍数据库中，与清代有关的占了很大一部分，如中国基本古籍库、中国方志库（共分五集，初集于2008年出版，收录省、府、厅、州、县2000种；二集于2013年出版，也收录方志2000种）、中国谱牒库（拟收录宋元明清历代家谱7000余种，年谱1200余种，仕谱300余种，日谱500余种，合计一万种，目前初集于2010年面世，共收录2000种）、中国类书库（包括《古今图书集成》）、中国俗文库、中国经典库、历代别集库、中国丛书库等，构成了一个庞大的数据库系统。由籍古轩研制的“中国数字方志库”是另一套大规模数字化综合地方志系统，涵盖了1949年以前约11000部志书，且含有一些极罕见的地理类著作，如晚清各省纂修大清会典后出版的省级舆图等。中国国家图书馆实施了“数字方志”工程，将该机构所藏数千部方志公开发布，扫描件比较清晰，但其界面及浏览速度、卷次著录等方面还存在较大改善空间。中研院以中科院北京天文台1985年编纂出版的《中国地方志联合目录》为基础，参照2000年新出版的《中国地方志总目提要》，建立了“中国大陆各省地方志书目查询系统”，基本涵盖宋元明清各时期及1949年后新编地方志目录^②。

清代报刊种类繁多，篇目数不胜数，利用起来非常不便，如《申报》，以往为了便于利用，不得不编制部头庞大的《〈申报〉索引》，但即使这样，也仅仅是报刊标题的索引而非内文，使用也十分不便。数字时代，报刊检索难的问题已迎刃而解。目前关于报刊的数据库已有数种，如爱如生的“中国近代报刊库”以“影响范围广、存续时间长、史料价值高”为遴选标准，从清道光十三年至民国年间的报刊类出版物中精选了3000种出版，全文有200亿文字之巨。此外，较常用的报刊数据库还包括“大成故纸堆”、《全国报刊索引》编辑部与上海图情信息有限公司研制的“晚清期刊全文数据库（1833—1911）”、“瀚堂近代报刊”等。另外，像《申报》、《东方杂志》等著名报刊也有几种专门的数据检索系统。

清史各专题数据库的建设也卓有成效，比较重要的有中研院近代史研究所的“清代粮价数据库”，这是中研院王业键院士自1970年代持续到2008年进行的清代粮价清单蒐集及粮价资料库建置的成果。该库以自乾隆元年开始各省向皇帝奏报的各省府、直隶州厅的主要粮食价格所形成“粮价清单”为基础，进行了数字化处理，可以系统展现粮价、人口等资料的时间与空间分布特性，目前，该数据库在清代经济史研究中发挥了巨大的作用。关于清代人口方面，有中研院

^① <http://hanchi.ihp.sinica.edu.tw/mql/login.html>，访问时间：2016-08-20。

^② <http://webgis.sinica.edu.tw/place/>，访问时间：2015-08-01。

“清代人口史研究资料库”、李中清、康文林基于八旗户口册与清代皇室族谱资料建设的中国多代人口系统数据库，包含辽宁、双城和皇族三个子数据库，其中辽宁、双城已免费公开^①。关于清代人物查询，有台北故宫的“清代档案人名权威档”，因该数据库是以军机处折件为基础制作的，具有较强的权威性，共约 6000 余人的信息，而且该人名数据库与台北故宫所藏的宫中档、录副奏折档建立了关联，功能十分强大。由哈佛大学燕京学社、中研院历史语言研究所、北京大学中国古代史研究中心合作开发的“中国历代人物传记资料库”（简称 CBDB）收录了近六万条清代人物数据，包括了籍贯、室名字号、著述等信息，并与地理信息系统实现了关联^②；李中清、康文林教授团队正在进行《缙绅录》数据库的建设工作，可望建立一套大规模的、分析型的清代职官信息数据。浙江省档案局利用了中国第一历史档案馆所藏的清代浙江官员引见档案建立了从康熙五十八年至宣统二年约两百年清代浙江官员履历表数据库，共 1145 条目录，每个条目由“姓名”、“籍贯”、“职官”、“年代”、“出处”五个字段组成^③。台湾大学数位人文研究中心建有“清季职官表查询系统”。^④关于清代司法方面，有南开大学中国社会史中心的“嘉庆朝刑科题本”资料库。

近年来，国家大型科技和基金项目也开始注重数据库的开发与建设，预计未来数年将会有新的多种数据库建成。

表 1 2009 年以来国家社科基金中与清史相关的数据库项目

| 年份 | 基金类别 | 项目名称 | 主持人 |
|------|----------|------------------------|-----|
| 2015 | 社科基金重大项目 | 清末民国社会调查数据库建设 | 黄兴涛 |
| 2013 | 社科基金重大项目 | 明清华北地区府县历史文化研究与专题数据库建设 | 戴建兵 |
| 2013 | 社科基金重大项目 | 清代灾荒纪年暨信息集成数据库建设 | 夏明方 |
| 2011 | 社科基金一般项目 | 锦屏文书数据库建设与村寨原地保护模式研究 | 曾羽 |
| 2009 | 社科基金一般项目 | 四库全书研究资源数据库建设 | 李芬林 |

来源：国家社科基金项目数据库，<http://fz.people.com.cn/skygb/sk/index.php/Index/seach>，访问时间：2016-09-01。

纵观清史数据库的发展，可以看出科研机构的推动与商业公司的运营是两大发展动力，前者利用自身积累的学术资源，借助政府科技计划的支持，发展出具有较高专业性的专题数据库，且不少面向学术界与公众免费开放；后者借助商业开发模式，大力发展各种规模较大的数据建设计划，如爱如生公司等开发的单个数据库，字数均可达数亿字之多，并借此形成了一定的盈利模式，数据库的规模迅速扩张，并占据了较大份额的数据库市场，具有了较大的议价能力，各高校图书馆数据库采购与续订的经费连年增长，已引起较大的反弹。^⑤

三

数据库的建设给清史研究带来的改变显而易见。清史与之前各断代研究的最大不同就在于资料的丰富程度非其他断代可比，故学界前辈常常提倡的“穷尽史料”的研究方法尽管仍是基本的指导性原则与追求的目标，但实际上对于清代这样一个时代而言，任何一个狭小的领域所涉及的资料已非单个研究者所能把握，更不要说整个清史领域了。因此，借助一定的手段，以较短的时间获取尽可能多的资料是清史研究工作者不得不面临的选择，数据库恰恰就提供了这种方便。

① <http://www.icpsr.umich.edu/icpsrweb/ICPSR/series/265>，访问时间：2015-08-01。

② <http://isites.harvard.edu/icb/icb.do?keyword=k35201>，访问时间：2015-08-01。

③ <http://www.zjda.gov.cn/dadb/dzda/qdda/>，访问时间：2016-09-01。

④ <http://ssop.digital.ntu.edu.tw/>，访问时间：2016-09-01。

⑤ 舒薇等：《高校图书馆数据库资源续订的调查与分析——以中南大学图书馆为例》，《学理论》2015年第16期；本刊编辑部：《学术数据库免费时代何时来临》，《科学导报》2016年第27期。

在研究方法上，大数据也将为清史研究带来一场全新的革命。以往学者曾经对史学研究中常用的一些逻辑方法进行了批评，李伯重以“宋代经济革命说”这一成说的形成为例，认为“选精”、“集萃”两种方法的错误在于将某些例证所反映的具体的和特殊的现象加以普遍化，从而使之丧失了真实性^①，实际上，“选精法”与“集萃法”在史学研究中是一种极为通行的做法。历史研究中，由于材料的缺失包括研究者个人收集资料的精力投入，“选精法”、“集萃法”在某些具体史学问题上仍不得不使用，但如果建立了较大规模的数据集，运用适当计量学的方法作为辅助，则上述方法论的缺陷也许可以得到一定程度的弥补，这也是数据库建设之于史学研究的最大意义，对于清史研究而言，尤其如此，利用数据会给清史研究带来一些全新的结论。这是许多倡导量化研究的学者一再强调的，讨论也比较多，此处不再赘述^②。

传统上由于资料获取便利程度的差异而带来的资料“鸿沟”是一个事实，资料保存地点进而发展成为重要的研究中心几乎是司空见惯的学术现象。清代中央一级的档案基本保存在北京和台北故宫，因此，在档案数据库建设之前，北京和台北的学者往往有“近水楼台”之便，外地学者阅读档案的时间成本和经济成本变得几乎难以忍受。即使是那些已经刊印的奏折资料，由于印数极少，刊印时间较早，不少学校并未购买。可以看到，过去许多清史论文不仅未刊档案利用得不多，就是已刊档案也很难讲得到充分利用，这极大降低了清史研究的水准。从理想状态而言，利用数据库系统，不仅使得身处不同地区的研究者获取资料和信息的速度大大加快，也有力地促进了海内外与国内各地区之间的学术交流，使得学术工作者在大致平等的资料面前展开自由竞争。尽管仍然存在资料获取便利程度上的差异，但无疑这种差异是可以被缩小的。当未来学者之间占有资料的差距缩小至最低限度时，学术研究模式将发生新的变化，过去靠占有独家资料从事学术研究的模式可能无法持续，这将对学者面对同一资料时提出问题的能力提出了更高的要求。

但同时，清史数据库也会是一把双刃剑，如处理欠妥，可能会造成新的资料鸿沟。目前，清史领域的数据库除了一部分是免费、公开资源外，不少重要数据的使用仍然受到较大的地域与人员限制，尤其是以档案数据库的使用最为显著。若干大型资料集由商业公司开发并按照商业模式运行，收费高昂，业已引起若干学校的反弹。可以想见的是，作为公益为主的数据库建设，如果受到不合理的费用与身份的限制，将会给一些非重点科研单位的研究者带来新的资料“鸿沟”。

在数据时代，史学中的考证学将迎来“春天”。黄一农在从事“红学”研究中，提出“e考据”的学术理念，并引起了学术界的争论。^③ 数据库尤其是各类史料的数据化，将使得对具体史学问题，尤其那些“关键词”较为确定的史事考证会非常有益，在史料利用范围的扩充上，信息时代的学人将具有天然的优势。事实上，我们已经看到最近若干年的学术论文在清史材料的使用上非常丰富，不仅常见史料娴熟使用，就是收藏于海外图书馆的文献及边边角角报刊上的信息都能够被采集到，新见资料层出不穷，因此，对清史工作者而言，今后的问题将不大可能是材料不足的难题，而是材料过多而亟需提高的史料剪裁能力。史料的易于搜寻与考证学的兴起也将带来新的问题，这便是已经引起学术界高度重视的史料的“肢解式”理解与“碎片化”问题在“数字化时代”未必会有改观^④，相反，还有可能会更加恶化，对史料的贯通理解与对宏观历史的把握仍将持续成为清史研究的难题。

① 参见李伯重：《“选精”、“集萃”与“宋代江南农业革命”——对传统经济史研究方法的检讨》，载《中国社会科学》2000年第1期。

② 梁晨、董浩、李中清：《量化数据库与历史研究》，《历史研究》2015年第2期等。陈志武等倡导推动的“量化历史讲习班”已举办四届，并创办《量化历史研究》辑刊，已出版两辑。

③ 谢乃和：《别让“e-考据”成为“伪考据”》，《中国社会科学报》2013年1月25日；张瑞龙《e考据是“立体”史学而非“伪考据”》，《中国社会科学报》2013年9月23日。

④ 可参一组“中国近代史研究中的‘碎片化’问题笔谈”论文，《近代史研究》2012年第4期、第5期。

四

清史数据库建设目前正呈现方兴未艾之势。可以看出，档案类文献的数据库由于收藏单位的集中，中央一级的档案中宫中档等基本整理完毕，目前只是有待扩大开放范围，吏科题本等其他档案也基本都已在中国第一历史档案馆的整理计划之中，按照目前数据库建设的成熟程度，预计会在不太长的时间内，将清代档案全部建设为数据库。至于地方上所藏的清代档案尤其是县级档案，如巴县档案、南部县档案等在国家清史工程的推动下，也取得了很大的成绩，这一领域的数据库建设预计将会在不不久的将来基本建设完毕。因此，未来清史数据库建设的一个方向更多将是专题数据库的建设。

但是，在清史数据库如火如荼建设的今天，我们更应该看到繁荣背后的危险与挑战。笔者在使用数据库的过程中，对于清史相关的数据库建设现状产生了以下一些思考，这些思考未必正确，只是一点浅见而已，敬请学界先进予以指正。

一、清史数据库迫切需要一个共同的数据库平台。目前，世界各国有关清史的数据库资源已经非常丰富，本文仅仅回顾了一些重要清史数据库，只是管中窥豹而已，实际上，与清史相关的各专题研究中，均有数量极其庞大的数据库资源，只是比较分散而已，如近代以来的图像资源，国内的首都图书馆等建立的有老照片、插图数据库，浙江图书馆编有“中国历代人物图像数据库”，国外如美国杜克大学公布的甘博有关中国的摄影、大英图书馆 Flickr 上的照片资源、东洋文库所藏关于中国的各类“写真”资料集等等。如能建立一个共同的数据库平台，将全世界的数据库资源陆续汇集起来，哪怕仅仅提供一些数据库的链接，建立起一个共同的检索目录，那也必将大大促进清史研究资源的共享。在这一方面，地理学数据平台的建设可以提供借鉴，如中科院建设的“地理空间数据云”、中科院地理所主持的“国家地球系统科学数据共享平台”等等，在整合国内外数据中心群、高等院校、科研院所以及科学家个人产生的数据资源，引进国际数据资源，接收国家重大科研项目产生的数据资源等方面，业已取得了很大的成绩。^①

二、需要强制性实施数据库信息公开。目前国内清史类的数据库建设得不少，从各种期刊的介绍中，可以看到各种与清史相关的数据库早已开始建设，但一直无法在网络上公开查到数据库的详情，要利用当然就更不可能。不少数据库的建设仅仅作为一个课题的副产品而存在，待课题结束，数据库的资料主要由一定圈子的学者分享。这对于推进整个清史领域的进步毫无益处。而且，数据库的研制往往是由各类课题作为支撑，严格意义上讲，它并非纯粹是个人科研的成果，而应该成为公共产品。因此，建议有关科研基金管理机构，出台相关规定，至少从受到国家经费支持的数据库项目开始实施数据强制公开办法，以此带动信息共享的学术理念在清史学界生根发芽。即使是部分数据库由于特殊情况，难以完全向公众开放，也应该本着便利的目的对学术界尽可能地开放，如关于清代档案这类对清史研究具有全局意义的数据库资源，应该早日制定开放的日程表。

三、清史数据库的建设可以尝试走公益与商业开发并行的道路。科研单位一般而言，建设史学数据库，多本着公益的目的，这当然应该鼓励。但数据库本身是一种高投入的行业，完全依赖于科研经费的投入并不现实，而且数据库开发以后，也面临后续更新、维护等问题，需要资金的持续投入。如能将数据库的建设进行适度的商业开发，以后续收益继续投入数据库的建设与维护中，相信会形成一种更好的良性循环，促进清史数据库建设的健康发展。事实上，目前学术界广泛使用的一些数据库多是商业开发的结果，如爱如生系列、文渊阁四库全书等。只是商业开发必须严格限定在一定范围内，是最低限度的商业化而不是完全以赢利为目的的商业化。

^① 该平台网址是：<http://www.geodata.cn/index.html>，访问时间：2016-09-01。

四、改进清史数据库重建设，轻关联的现状。目前，国内数据库的开发专题性强，在设计数据库时，基本都考虑到该数据库的个性化特色，而较少考虑到未来数据库之间的关联问题，因此，目前清史数据库虽多，但其功能似乎还未完全发挥出来。在未来数据库建设过程中，需要进一步加强数据库之间关联的可能性的考量，这就需要在数据库建设中，尽快制定出一套标准的适合清史学科的技术规范。

五、在加强清史数据库建设过程中，要更加重视传统史学的训练。数据库的使用为史学研究提供了极大的便利，尤其是对于史料的收集而言，使得新一代史学工作者的技术优势显露无疑。但毋庸讳言的是，数据库的使用也带来了一些弊端，这就是在写作史学论文时，习惯于用关键词检索的方式获取史料，而非通过深度阅读；引用古籍常常以电子资料为基础而失于与权威文本核验，造成了史料引用上的不严谨现象十分突出；对于史料的理解常常是断章取义，而未能对文本前后语境有所体认。况且，史学研究中，关键词的设定不会自动生成研究论题的所有材料而只是部分材料，关键词的设定限制了自由思考的空间，一定程度上是根据问题找相应材料，陷入“史以填论”的陷阱；况且，以清史资料之多，在数据库大量涌现的大背景下，任何一个论题的研究都难免陷入浩如烟海的史料之中而疏于择取，故有的论文往往史料十分详赡却不忍卒读。应该说，无论未来数据库发展到何种程度，它始终是一个辅助手段而非主要手段，对史料的深度阅读永远不可替代，传统史学的一套训练方法始终是史学的最重要、最基本的训练，始终富有生命力，在数字时代，它不是不重要了，相反，应该是更加重要了。只有将传统史学的严谨训练融入到现代技术手段所引发的史料爆炸浪潮中，清史研究才能真正迎来新的“春天”。

References

Mao, Jianjun. *Theory and Practice of Ancient Books Digitization*. Beijing: Aviation Industry Press, 2009.

Hsiang, Jie ed. *Digital Humanities and Historical Research*. Taipei: National Taiwan University Press, 2011.

Li, Bozhong. “‘Choosing the Refined’ and ‘Collecting the Essential’ and the Agricultural Revolution South of the Yangtze River During the Song Dynasty—A Criticism of Traditional Methods in the Research on Economic History.” *Social Sciences in China* 1 (2000): 177–208.

Jin, Guantao and Liu Qingfeng. “The Truth of History: Database Applied in the Historical Research.” *Studies in Qing History* 1 (2008): 90–108.

Liang, Chen, Dong Hao and Li Zhongqing. “Quantitative Database and Historical Research.” *Historical Research* 2 (2015): 113–128.

Xia, Mingfang. “Big Data and Ecological History: The Compilation and Database Construction of the Historical Sources of Chinese Disaster History in an Information Age.” *The Qing History journal* 2 (2015): 67–82.

Shu, Wei, Zhou Fen and Li Liangxian, eds. “The Investigation and Analysis of the Renewel of the Database in University Library: in the case of Central South University Library.” *Theory Research* 16 (2015): 209–210.

The Establishment of Databases and Research in Qing History

HU Heng (The institute of Qing History, Renmin University of China; huhengqss@ruc.edu.cn)

There have been breakthrough achievements in the building of historical databases for Qing history, and the basic documentary databases and specialized databases have developed together. The establishment of Qing history databases and the subsequent quantification of historical research has overcome the lack of systemic statistical data in traditional historical research and other serious deficiencies in statistical research. These databases have greatly improved the opportunities for using quantitative methods but there are also new issues in the use and understanding of historical materials. The development of local historical statistical archives requires a parallel development of platforms for sharing data and the balancing of public benefits and commercialization. At the same time, in the age of digitization it is still important to provide training in traditional historical methods.