

BIG

DASHUJU SHIDAI DE SHILIAO YU SHIXUE

《史学月刊》编辑部 编

大数据时代的 史料与史学

DATA



人民出版社

DASHUJU SHIDAI DE SHILIAO YU SHIXUE

大数据时代的 史料与史学

《史学月刊》编辑部 编



人民出版社

责任编辑:郭 娜

封面设计:周方亚

图书在版编目(CIP)数据

大数据时代的史料与史学/《史学月刊》编辑部 编. —北京:人民出版社,
2017.11

ISBN 978-7-01-018060-1

I. ①大… II. ①史… III. ①数字技术-应用-史学-研究 IV. ①K03-39

中国版本图书馆 CIP 数据核字(2017)第 200941 号

大数据时代的史料与史学

DASHUJU SHIDAI DE SHILIAO YU SHIXUE

《史学月刊》编辑部 编

人民出版社 出版发行
(100706 北京市东城区隆福寺街 99 号)

北京市文林印务有限公司印刷 新华书店经销

2017 年 11 月第 1 版 2017 年 11 月北京第 1 次印刷

开本:710 毫米×1000 毫米 1/16 印张:28

字数:360 千字

ISBN 978-7-01-018060-1 定价:72.00 元

邮购地址 100706 北京市东城区隆福寺街 99 号

人民东方图书销售中心 电话 (010)65250042 65289539

版权所有·侵权必究

凡购买本社图书,如有印制质量问题,我社负责调换。

服务电话:(010)65250042

目 录

历史学:当小时代遇见大数据	1
---------------------	---

第一篇 大数据时代的史学发展趋势

论互联网时代的历史学	7
大数据时代史学的多元发展趋势	46
当记忆成为常态,历史学何去何从	51
大数据时代史学研究的趋势与反思	66
大数据会对中国史学产生根本冲击吗	82
大数据时代的人文关怀	
——浅议史学方法革新与史学观念的互动	94
大数据视野下历史气候研究的史料与史学	
——以游牧民族迁移及王朝兴亡为例	108
大数据时代“全史料数据平台”构想与建设	
——以日本现有历史资料数据库为例	135

第二篇 数据资料搜集利用的经验与建设

历史学数字资源利用的误区及其应对	149
数据库与地方党史文献的发掘及利用	158
中国古史研究中古籍、学术论著数字化的应用与思考	180
西方学界蒋介石研究的范式转移与取向变迁 ——以英文论著为中心的探讨	195
以罗斯福总统图书馆为例谈历史文献数据库的运用	226
国内现存滇西抗战口述文献综述	234

第三篇 大数据时代史学研究方法的借鉴与革新

《全宋文》北宋作者的时空分布特征 ——基于 QGIS 等技术的研究	251
大数据时代的珍稀史料考释及史学价值	267
数字化技术在文博考古中的应用	287
碑刻中的工价资料的处理与应用 ——以大数据为研究工具的尝试	302
“数字人文热”背景下的冷思考	318

第四篇 大数据时代史学研究的拓展与开新

大数据时代的中国慈善史研究及其运用	327
大数据与国内美国史研究的机遇与挑战	339
地狭人稠型城市的困境与治理 ——以港九时期(1860—1897)的香港为例	361
史料的形态与历史的书写:以德国中世纪晚期和近代早期的 历史研究为例	392
知识即权力? 英国医生的话语建构与发展趋势(1800—1948)	408
大数据时代的史学因应 ——“新史学青年论坛:大数据时代的史料与史学”会议综述	427

历史学：当小时代遇见大数据^{*}

张宝明

众所周知，历史学是一门古老的学科，而历史学研究的厚积薄发、薪火相传则在于青年人才的发现与成长。本来，从人文学科的视角来说，无论是文化还是思想，抑或学术，其间的关系并不存在新、旧的截然判定。说穿了，现代与传统的环环相扣、绵延接续在历史学这里表现得尤为突出。因此，我们有理由对克罗齐“一切历史都是当代史”^①的判断做出“于我心有戚戚焉”^②的回应。从这个意义上说“新史学”，无疑是指其新时代、新方法之“新”。的确，我们处于“小时代”的背景，这是与时俱进的机遇之“新”；与此同时，我们又赶上了“大数据”，这是史无前例的挑战之“新”。及此，我们就分别说说千年不遇的机遇与挑战。对这双重之“新”，我们可以用“一则以喜一则以忧”来表达这个“参半”的情境。

一、喜：历史写作人皆可为

史学研究的基础在于掌握丰富的史料。依据充足的史料进行科学分

^{*} 2016年11月3—4日，“首届新史学青年论坛：大数据时代的史料与史学”在天津南开大学召开，该文为河南大学副校长张宝明在开幕式上的发言。（张秀丽 整理）

^① [意] 克罗齐：《历史学的理论和实际》，商务印书馆 1982 年版，第 2 页。

^② 《孟子·梁惠王上》。

析论证并由此得出自圆其说的结论。可以这样说,搜集和积累史料是史学研究者的基本功与天职。在大数据时代以前,史学研究者做学问、爬梳史料的功夫,主要是做笔记和抄卡片,并进行整理、分类,这是历史学者的基本功,也是一个能够补拙的笨功夫。正是这个笨拙的基本功成就了一批又一批卓有成就的历史学家。陈登原教授曾经教导学生以卡片战术治学的事迹不是一个传说。老先生不但言传身教,而且是亲力亲为的“夫子自道”。他每天抄写若干卡片,集数十年之功而笔耕不辍,终成为一代大师的例子,为传统史学家的方法和路径做了立体而生动的注脚。

进入大数据时代,互联网、多媒体以及大型数据库的建设为获取史料提供了数字化、信息化、快捷化的检索手段。随着信息平台的多样化和史料查阅的公开化,历史研究随之呈现出民主化、公平化、多样化的特征。这一改此前资料垄断、即使同行也如隔山的特点。历史学研究的门槛在降低,历史学家的权威被解构,历史写作人皆可为。的确,史料检索功能的高效快捷,可以在短时间内获取过去阅读、摘抄方法上难以想象的信息和数据,研究者的选题眼界大为开阔,研究方法也在更新。当今史学界的年轻学人运用电子文献资源查找资料已经十分普遍,都是大数据的受益者。他们的获得感使其离成为历史学家的梦想又近了一步。

二、忧:如何不被新技术绑架

不过,我们在为历史学研究的机遇欢呼雀跃之际,也要清醒地看到与之俱来的挑战。这即是我们所说的“一则以忧”。

首先是如何在信息爆炸、数据叠加的时代不为新技术、新方法、新手段绑架而不能自拔。也就是说,面对海量的资料、无限的信息,如何科学鉴别、筛选、分析与利用它们?著名的历史学家都具有训练有素的“内功”。这个内功指的是内化于心的功底——史才和素养,而不是外在于

人、堆积如山的底料。最基本的史学素养,必须通过对最基础的原始资料的阅读来获得。借助先进的电脑网络手段,史学研究者可以快速、便捷地检索到大量史料,由于受到知识结构、学术积累和理论修养的局限,我们无法确保自己能够准确分析和合理运用史料。在大数据时代,史料处理的功夫在于史料之外的史才与素养。进一步说,这个史才与素养也就是胡适所说的“眼光、识力与笔力”^①。

其次是如何处理读书与搜集史料的关系。作为一门古老而健全的学科,历史学研究有全面系统、相对成熟的学术体系。在著书立说的过程中,它要求点、线结合,纵、横结合,史、论结合,从而在整体意识上编织出繁花似锦的历史学网格。无论是法国年鉴学派的“长时段”理论,还是马克思主义的历史“阶段论”,都意在强调从大局、全局乃至整体出发去理解、同情乃至想象,以尽可能精准地认识和分析问题。由此,才会有建立在感性认识上的解读和高屋建瓴、鞭辟入里的论析。在大数据时代以前,读书是搜集史料的前提。不读书,根本不知道也无从感知,更无从深刻地理解。如果以检索代替读书,用关键词搜取代资料查阅,用碎片化的信息概括全面、系统的内容,那样的文章即使再像模像样,也只能是一堆用毫无生气的材料拼凑起来的“编年史”,而非历史学家经过大脑思考、透过心灵体验造就的“真历史”^②。李根蟠先生就曾指出:“不管电脑检索如何方便,它代替不了读书。”^③

进一步说,史料数据化虽然具有利用便捷的优点和广阔的发展前景,但不能改变学术研究的本质,它仍然仅仅提供了研究的史料基础,史学成就的取得还是依赖史学界诸多学者的思考、探索和相互讨论。以今天历史研究的情形而言,无论是选题还是结论都是经过历史学家或历史工作

① 胡适:《历史的文学观念论》,《新青年》第3卷第3号,1917年5月1日。

② [意]克罗齐:《历史学的理论和实际》,商务印书馆1982年版。

③ 李根蟠:《历史学习与研究方法漫谈》,《中国农史》2004年第2期。

者精心、审慎选择与多层过滤、反复思考的结果,这在很大意义上呈现为当下与过去的相似性、亲近性与有机性。死的历史只会遭到静卧、尘封的待遇,只有活的历史才会得到激活、营造的礼遇和待见。我认为,历史学家尤其是思想史家,理应“在死亡的历史废墟上营造一方生命的绿洲”^①即是源于这样的一个感同身受。进一步解读这句话,也可以说是以当下活生生的心灵去观照并去激活那死亡的废墟。无论是历史学家还是考古学家,如果他苦心孤诣、殚精竭虑,甚至皓首穷经、死而后已地抓住“孔子的第几个脚指头有脚气”之类的问题猛捧,这不只是历史学家的悲剧,乃是关乎这个学科是“to be or not to be”的闹剧。何以故?因为那些问题并不关乎现实与当下,即使是铁证如山的鸿篇巨制,也不过是轻如鸿毛的过眼云烟。

三、元命题:小时代与大数据

“喜忧参半”之后,我们还是要回到并面对我们的元命题:小时代与大数据。也许,小时代并不是我们的选择,我们只能接受、顺应或者说面对、经营这个时代。面对大数据,如何变被动为主动,将失控、漂浮的无限量信息尽在掌控之中,才是历史学家的不二选择:一方面摆脱随波逐流的漩涡而为我所用,另一方面又不失时机地寻求事半功倍的效果而不失人文主体性。一言蔽之,小时代,我们应该努力让这个“小”榨出“大”来,以“小”见“大”;大数据,我们则应该以不怨天不尤人的心态从容“爆料”,以“大”压“小”。或许,正是由大、小互动而生的厚薄、轻重之博弈,才会有效地规避浅薄、失重。由此历史学与历史学家才不会辜负这个时代的担当。

^① 张宝明:《启蒙与革命——“五四”激进派的两难》“导言”,学林出版社1998年版,第2页。

论互联网时代的历史学

第 一 篇

大数据时代的史学发展趋势

论互联网时代的历史学

大数据时代史学的多元发展趋势

当记忆成为常态,历史学何去何从

大数据时代史学研究的趋势与反思

大数据会对中国史学产生根本冲击吗

大数据时代的人文关怀

——浅议史学方法革新与史学观念的互动

大数据视野下历史气候研究的史料与史学

——以游牧民族迁移及王朝兴亡为例

大数据时代“全史料数据平台”构想与建设

——以日本现有历史资料数据库为例

BIG DATA

论互联网时代的历史学

李振宏*

两年前,《史学月刊》曾经编发过一组题为《计算机技术与史学研究形态》的笔谈文章,我在给这组文章写的“编者按”中说:“最近30年来,随着计算机技术的广泛应用以及在人文社科研究领域的普及,学术研究的路径、手段以至研究形态,都在发生着深刻的变化,一向被认为是最古老最传统的历史学科也不例外。如今的历史学研究,不光是搜集资料的方式或手段由于电子文献检索的推广而变得极为便捷,以往竭泽而渔、将材料一网打尽的梦想轻而易举地变为现实,而且数据库方法也渗透到逻辑分析的领域,影响到学者思维方式的改变。古老的历史学科,似乎正在展开一个学术科学化的发展方向。但也不容否认,计算机技术的应用也带来不少新的问题,诸如以材料检索代替读书的便捷路径是否可靠,读书与检索的关系如何处理,数据分析与逻辑分析、历史分析的关系,如何看待数据库方法的局限性,在数据库方法日益推广的条件下理论思维的价值和意义,如何估价计算机技术带来的研究形态的变化,在新的科研条件下新一代史学研究人才如何培养,等等,都是需要认真面对和慎

* 李振宏,河南大学黄河文明协同创新中心、史学理论与史学史研究所教授。

重思考的问题。”^①所以会组织这样一组笔谈文章,就是我们深深地感受到了在计算机网络技术日益普及的今天,传统历史学正在发生着一场深刻的变革,如何正确地面对这场变革,是当代史学发展需要回答的一个重要问题^②。

关于这场史学变革,人们都惊呼进入了一个新的学术时代,并对之有各种称谓,诸如数字化时代、数据库时代、计算机时代、互联网时代、信息化时代,等等。总之,这是一个电子、信息、人工智能越来越深化发展,越来越改变着人们的生活方式、思维方式并影响到学术研究形态深刻变化的时代。本文将直面这个学术时代的种种变化,以期为我们更好地应对这场变革提供一些初步的思考,希望听到学界同人校正和批评的声音。

一、当前学界关于史学变革的认识

关于互联网、大数据、传统文献数字化所带来的史学变革,人们已经发表了不少看法,这些看法构成本文讨论的基础和前提。所以有必要对学界已有的研究给予介绍和评述。

① 《计算机技术与史学研究形态》“编者按”,《史学月刊》2015年第1期。本期笔谈共发表论文五篇,分别是:乔治忠:《历史研究电子资源运用的兴利除弊》;王子今:《“史识”与计算机“利器”》;王文涛:《信息时代的文献阅读和史料检索》;陈爽:《回归传统:浅谈数字化时代的史料处理与运用》;周祥森:《空间转向:电子传媒技术与当代史学形态》。

② 事实上,关于史学研究形态的变革及其应对,史学界在很早以前就已经开始关注了。除了所谈《史学月刊》组织的笔谈之外,《史学月刊》2001年就开始增设“电脑与史学应用”栏目,积极组织编发这方面的论文;2011年第4期《史学理论研究》杂志,编发了一组六篇《互联网与史学观念变革》的笔谈文章;2015年12月,上海大学举办了以“大数据时代下的历史研究”为议题的国际学术研讨会;最近《史学月刊》与南开大学历史学院联合举办的“新史学青年论坛:大数据时代的史料与史学学术研讨会”,于2016年11月在南开大学举行;最近几年来,基于互联网的“e-考据”的争论也方兴未艾。凡此种种,都说明由互联网、大数据、传统文献数字化所带来的史学形态变化问题,已经越来越紧迫地摆在我们面前。

1. 从资料获取方式方面所作的评论

在以往的研究中,人们大多都注意到了传统文献数字化和大数据、互联网技术所带来的研究手段的改变,特别是获取资料方式的改变,盛赞当今人们实现了传统史家将史料一网打尽、竭泽而渔的梦想。文献检索的便捷,网络资料的丰富性,文献资源的共享,极大地改变了史学研究的效率和方式。这既是最表面化的表现,最易观察的方面,同时也是传统文献数字化给人们留下的最初印记,是每个学者都有的切身感受。譬如,陈爽写道:

30年间,古代历史典籍的载体大多已经由纸质书本转换为电脑可识别的文字符号,新一代学者对电脑写作和网络查询早已驾轻就熟,数字化检索已成为史学研究必不可少的辅助手段,史家对于史料的获取方式经历了亘古未有的革命。数字媒介的引入,促成了阅读方式和写作方式的改变,也使知识发现途径与处理手段都发生了变革。甚至有学者十分乐观地断言“一个有机会孕育‘e-考据学派’的时代或已出现”。^①

陈春声发文赞誉新时代的历史学变革说:

传统时代的历史学者皓首穷经,有时可依赖对冷僻资料的占有、对新资料的发现、对浩瀚文献中某个词句的挖掘或解读而对学术有所贡献。然而,进入21世纪之后,占有所谓冷僻资料或发现新资料这类具有“学术积累”意义的工作,已经越来越成为普通史学工作者

^① 陈爽:《回归传统:浅谈数字化时代的史料处理与运用》,《史学月刊》2015年第1期,第14页。

日常研究过程的一部分,毫无惊喜可言。更为重要的是,在“数字人文”的时代,由于“数字仓储”和“数字图书馆”的大量存在和在互联网上的利用便利,由于海量的资料文献可以“全文检索”之类的方式便利地查询,由于“文本数字挖掘”蕴含着几乎没有限制的“创造”与“重构”史料的可能性,传统条件下一位学者需要花费数月、数年光阴,甚至要花费毕生精力进行比对、校勘、辑佚、考订才得以解决的问题,现在可能在计算机网络上花费数秒钟、数分钟就可以有相当确切的结果;而原来因为缺乏史料,许多传统历史学家认为不能研究的重要问题,在“数字人文”的背景下,变得有点“唾手可得”。^①

姜义华撰文《大数据催生史学大变革》,也主要是强调由于历史资料利用方式改变所带来的变革。他说:

大数据使历史资料利用产生革命性变革……大数据具有速率、体量、多样性、价值密度四大特征。在历史学领域,大数据成百倍、成千倍地扩大了历史资料的范围。尤其突出的是,除传统的文献与遗址、遗物外,人类存留的所有纸质的、音像的和其他物质的、非物质的资料,几乎都可被用来协助复原历史的本来面貌……大数据通过数据采集、数据储存、数据分析、数据管理为大范围、中时段、长时段定量分析提供了现实的可能。历史学研究与大规模基因检测相结合,有力推动了家族史、宗族史、移民史、民族史研究,就是一个成功的实践。大数据使历史资料利用产生革命性变革,更集中表现在大数据有助于人们获得新的认知,因为它可以较好地用数字模拟方式在一

① 陈春声:《新一代史学家应更关注“出思想”》,《史学月刊》2016年第6期,第16—17页。

个大范围内展现历史场景。大数据信息系统还可以较方便地将历史活动中各种要素全面、综合、动态地展示出来,这样,它就可能为历史研究创造新的价值基准^①。

最近一些年,台湾学者黄一农提出一个著名的“e-考据”概念,即基于互联网、电子资源所进行的历史考据。“e-考据”说引起了一些讨论,不时有相关论文问世,而究其实,所谓“e-考据”,也是从文献利用的角度对新形态史学研究方法的一种表述。一篇介绍黄一农“e-考据”的文章说:

1994年,黄一农曾经以利玛窦在中国最亲密的友人瞿汝夔为素材写过一篇3页的《瞿汝夔家世与生平考》,十年后,当他准备将该文收为其研究中国天主教史的新书《两头蛇》中的一章时,黄一农选择了重写……通过台北汉学研究中心所制作的《明人文集联合目录及篇目索引资料库》,黄一农从其中收录的2500多种明人文集中检索到了丰富到难以令人相信的素材。通过这些素材,黄一农不仅可以了解到瞿汝夔家族中其他人的资料,甚至还可以知道其研究对象的外貌、身高、性格及其人脉关系,并有条件探索瞿汝夔与嫂嫂发生暧昧关系的背景环境,而这之前仅仅有三四本文集可以参考。通过运用这些资料,黄一农得以像写侦探小说一样细致而逻辑严丝合缝地还原历史现场,这一次,他写了30页,并且在《两头蛇》一书中严肃地提出了“e-考据时代”这一新的概念。

70多年前傅斯年先生大力提倡以科学方法治史,但是以其为代表的史料学派,无可否认尚未能令中国史学界的治学方法普遍发生

^① 姜义华:《大数据催生史学大变革》,《中国社会科学报》2015年4月29日。