

# 唐宋历史评论

第三辑

中国人民大学唐宋史研究中心

包伟民 刘后滨 主编

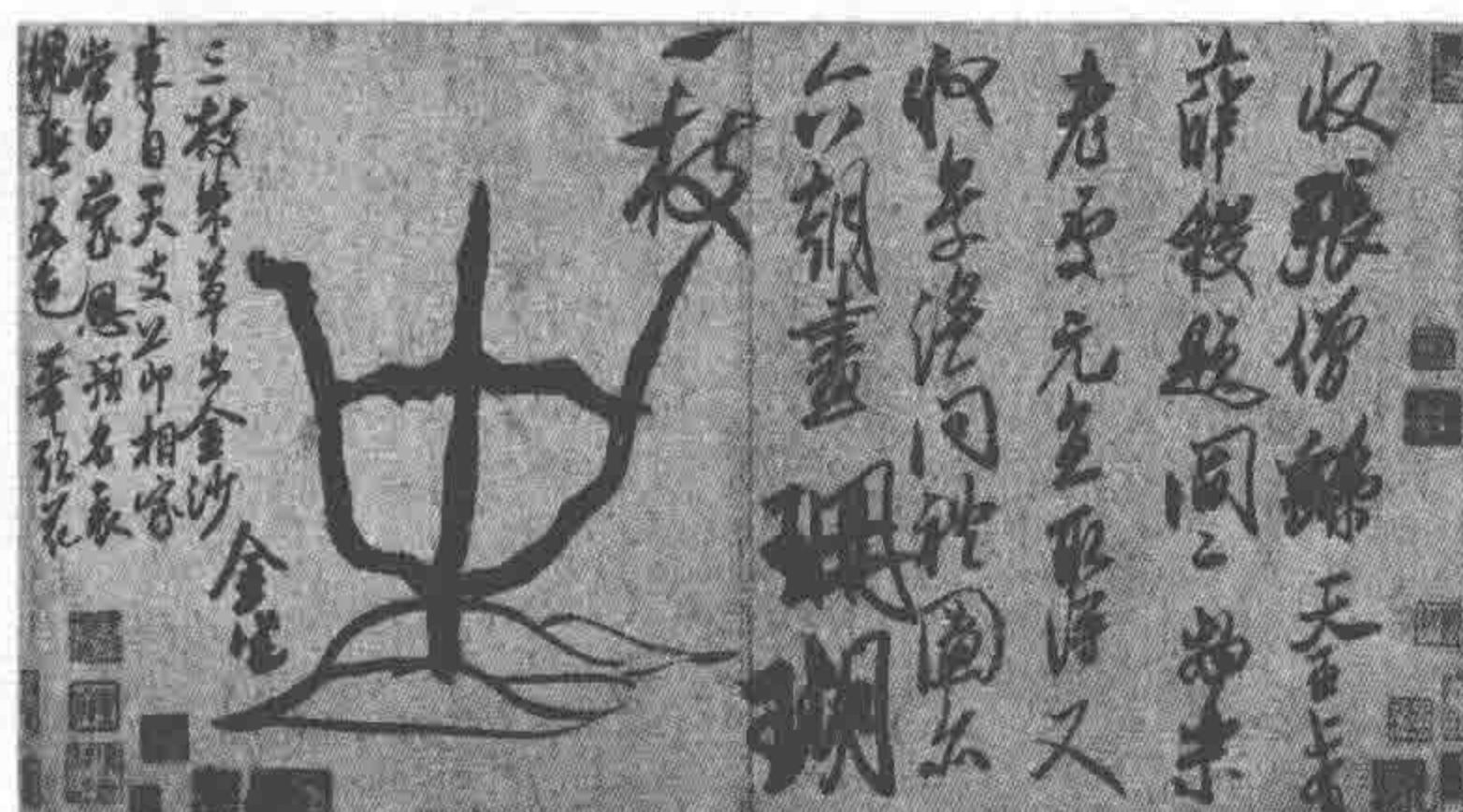


- |                         |     |
|-------------------------|-----|
| 唐宋史研究中的数字化语文学           | 魏希德 |
| 论唐代皇帝与高级官员政务沟通方式的制度性调整  | 叶 炜 |
| 唐前期外官月料分配比例考释           | 朱博宇 |
| 唐宋蒙学书系年考证与研究            | 李裕民 |
| 《宋史》的史源及其相关问题           | 顾宏义 |
| 试论宋代地方治安维护体系中的巡检        | 赵冬梅 |
| 文书、信息与权力：唐代中枢政务运行机制研究反思 | 刘后滨 |
| 北宋三司研究述评                | 张亦冰 |

# 唐宋历史评论

第三辑

中国人民大学唐宋史研究中心



包伟民 刘后滨 主 编  
李全德 王 静 执行编辑

## 图书在版编目(CIP)数据

唐宋历史评论·第三辑 / 包伟民, 刘后滨主编. --  
北京 : 社会科学文献出版社, 2017. 6  
ISBN 978 - 7 - 5201 - 1017 - 4

I. ①唐… II. ①包… ②刘… III. ①中国历史 - 研究 - 唐宋时期 IV. ①K240. 7

中国版本图书馆 CIP 数据核字 (2017) 第 148918 号

## 唐宋历史评论(第三辑)

主 编 / 包伟民 刘后滨

出 版 人 / 谢寿光

项目统筹 / 宋月华 李建廷

责任编辑 / 周志宽 胡百涛

出 版 / 社会科学文献出版社 · 人文分社 (010) 59367215

地址：北京市北三环中路甲 29 号院华龙大厦 邮编：100029

网址：[www.ssap.com.cn](http://www.ssap.com.cn)

发 行 / 市场营销中心 (010) 59367081 59367018

印 装 / 三河市东方印刷有限公司

规 格 / 开 本：787mm × 1092mm 1/16

印 张：24.25 字 数：385 千字

版 次 / 2017 年 6 月第 1 版 2017 年 6 月第 1 次印刷

书 号 / ISBN 978 - 7 - 5201 - 1017 - 4

定 价 / 69.00 元

本书如有印装质量问题, 请与读者服务中心 (010 - 59367028) 联系

 版权所有 翻印必究

# 目 录

---



## · 笔谈 ·

- 唐宋史研究中的数字化语文学 ..... 魏希德 撰 徐力恒 译 / 3  
唐代人物资料的数据化：中国历代人物传记资料库（CBDB）近年  
工作管窥 ..... 徐力恒 / 20  
家族、婚姻与道学：《仙溪志·人物传》中的社会  
关系 ..... 李宗翰 郑 莉 / 33

## · 专论 ·

- 论唐代皇帝与高级官员政务沟通方式的制度性调整 ..... 叶 炜 / 49  
唐前期外官月料分配比例考释 ..... 朱博宇 / 73  
穆格山粟特文婚约译注 ..... 张小贵 庞晓林 / 107  
唐宋蒙学书系年考证与研究 ..... 李裕民 / 126  
《宋史》的史源及其相关问题 ..... 顾宏义 / 161  
试论宋代地方治安维护体系中的巡检 ..... 赵冬梅 / 189  
明代方志复原宋元地方祠庙体系可能性探讨  
——以《弘治温州府志》祠庙记录为例 ..... 吴铮强 / 211

• 札记 •

唐琐事杂考 ..... 梁太济 / 231

历史学如何算起来？

——从北宋耕地面积、粮食亩产量等数字说起 ..... 张邦炜 / 243

宋本《洪氏集验方》印纸属地的确认及其意义 ..... 孙继民 / 252

• 述论 •

文书、信息与权力：唐代中枢政务运行机制研究反思 ..... 刘后滨 / 265

北宋三司研究述评 ..... 张亦冰 / 288

• 书评 •

陆扬《清流文化与唐帝国》 ..... 李鸿宾 / 323

陆扬《清流文化与唐帝国》 ..... 陈志坚 / 334

李碧妍《危机与重构——唐帝国及其地方诸侯》 ..... 胡耀飞 / 350

柏文莉《妓、妾与贞节观：1000~1400年中国的性别与社会

转变》 ..... 王申 / 366

方诚峰《北宋晚期的政治体制与政治文化》 ..... 杭州宋史论坛 / 373

# Contents

## Forum

The Uses of Digital Philology in Tang – Song History

Written by Hilde De Weerdt, Translated by Lik Hang Tsui /3

New Approaches toDigitizing Tang Biographical Data for the China

Biographical Database Lik Hang Tsui /20

Lineages, Marriages and *Daoxue*: Social Relations in the Biography

of *Xianxi* Gazetteer Lee Tsong – han , Zheng Li /33

## Research Articles

On the Changing Methods of the Institutional Communication between the Emperor  
and Officials in the Central Government in the Tang Dynasty Ye Wei / 49

A Study on the Distribution Percentage of the Monthly Salary Payments of Local  
Officials in Early Tang period Zhu Boyu / 73

An Annotated Translation of the Sogdian Marriage Contracts in the Documents of  
Mount Mugh Zhang Xiaogui , Pang Xiaolin / 107

On the Chronological Dating and Study of the Primers in the Tang and Song  
Dynasties Li Yumin / 126

The Original Sources for *Song Shi* and the Related Issues Gu Hongyi / 161

A Preliminary Study on the Xunjian in the Public OrderMaintaining System of  
the Song Dynasty , 960 – 1279 Zhao Dongmei / 189

On the Possibility of Reconstructing the Local Temple and Shrine Systems in the  
Song – Yuan Periods from the Ming Gazetteers: A Case Study on the Temple

- Records in the *Record of the Wenzhou Prefecture* in the Reign Period  
of Hongzhi Wu Zhengqiang / 211

### Notes and Essays

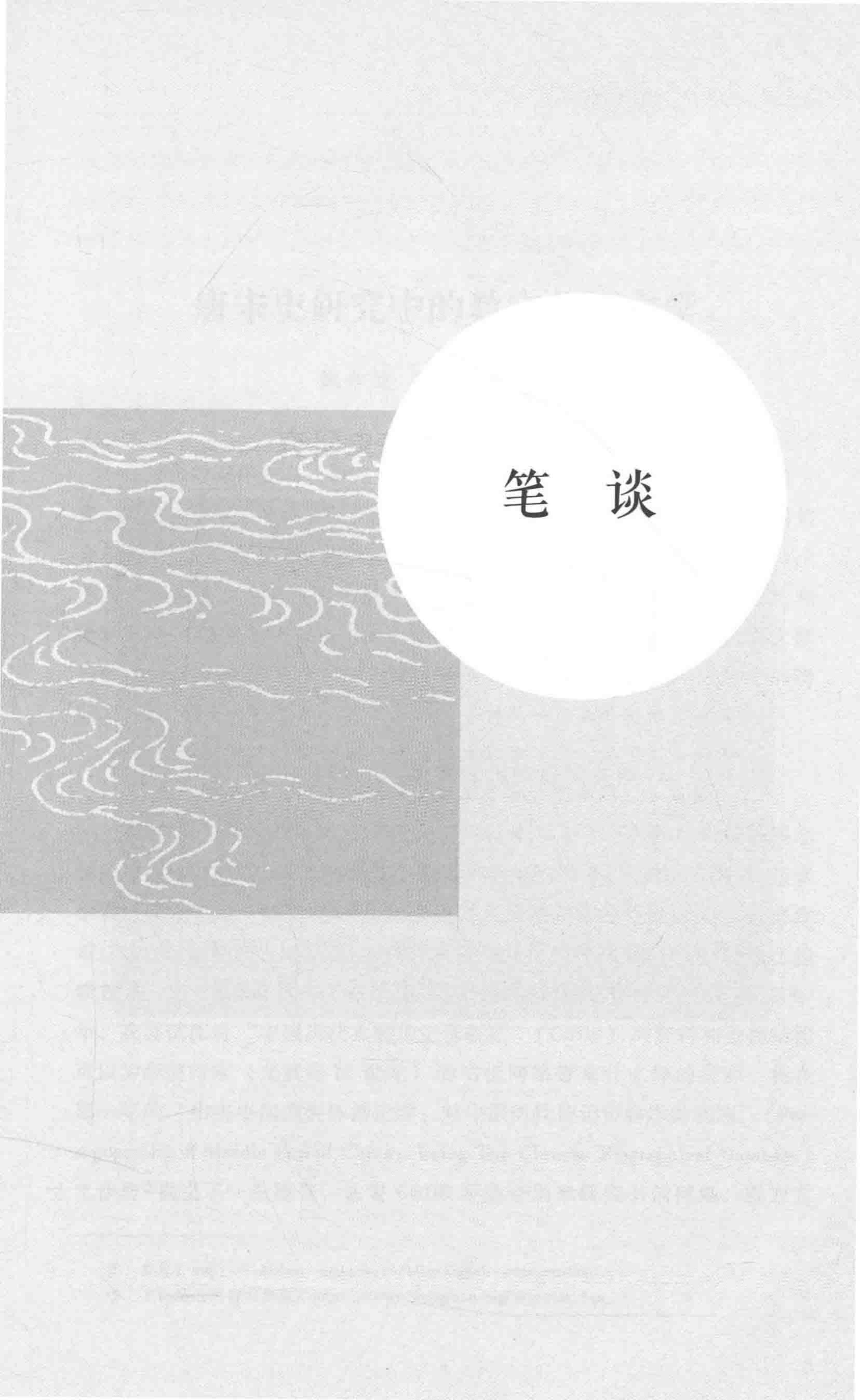
- Miscellaneous Notes on the Tang History Liang Taiji / 231
- How to Measure in History? ——On the Numbers of the Arable Land Size and  
the Yield per *mu* unit Zhang Bangwei / 243
- The Verification of the Original Place for the Printing Paper of the Collected  
Verified Recipes of the Hong Family and Its Implications Sun Jimin / 252

### Review Articles

- Documents, Information, and Power: A Reflexive Thinking about the Study  
on the Dynamics of Political Governance in the Central Court of the Tang  
Dynasty Liu Houbin / 265
- An Overview of the Studies on Three Central Political Institutions in the  
Northern Song Dynasty Zhang Yibing / 288

### Book Reviews

- Lu Yang, *Literocracy and Empire: A Study of Political Culture of the  
Tang Dynasty* Li Hongbin / 323
- Lu Yang, *Literocracy and Empire: A Study of Political Culture of the Tang  
Dynasty* Chen Zhijian / 334
- Li Biyan, *Crisis and Restructure: The Tang Empire and Its Regional Powers*  
Hu Yaofei / 350
- Bossler, Beverly, *Courtesans, Concubines, and the Cult of Female Fidelity:  
Gender and Social Change in China, 1000 – 1400* Wang Shen / 366
- Fang Chengfeng, *Political Institution and Political Culture in Late Northern  
Song* The Song Forum in Hangzhou / 373



# 笔 谈

## 数字人文与唐宋史研究

### 编者按语：

作为在唐宋史研究领域发展起来的数据库，哈佛大学、台湾中研院和北京大学联合开发的“中国历代人物传记资料库”(CBDB)项目已有十多年了。项目组在开发工作中积累了一些经验，唐宋史学者利用该资料库也做出了一些可喜的成果，并且借助其数据开发了一些数字化研究工具。为了分析介绍这几方面的最新进展，反思唐宋史的数字人文研究，本辑组织刊发了三篇稿子。魏希德回顾了她涉足数字人文研究的历程，着重探讨新型数字化研究工具为唐宋史研究带来的新视角。徐力恒的文章分析了CBDB项目近年把唐代人物资料转化为历史数据的工作，其中反映了学界当前建设史学数据库的新方法与所面临的挑战。李宗翰和郑莉的论文则提供了利用CBDB数据研究地方人物关系的案例，并藉此对这种研究方法做出反思。

# 唐宋史研究中的数字化语文学

魏希德 撰 徐力恒 译

在过去的五年左右，研究中国史的学者和研究其他地区的历史学者一样，意识到数字史学并非只是把文本变成量化资料，或只是把这类资料整合起来。<sup>①</sup> 不少人发现，历史学者利用数字化语文学（digital philology）的方法可以辅助并丰富他们在研究文本和物质文化中使用的分析方法和诠释策略。在这篇短文里，我将概括讨论我接触到的一些用于唐宋文献的数字化研究方法，并讨论我如何将之整合到唐宋政治史和思想史的研究之中。

## 一 从独立的数据库到数据的关联

大部分历史学者已经接触数字化技术十到二十年了。除了最容易联想到的用计算机打字以外，还有为了搜集资料进行简单的检索。此外，很多学者已经在使用一些常见的办公软件，用来创建表格、统计表、数据库等。这些做法逐渐在辅助或已然取代了过去的读书笔记和卡片。我自己探索数字人文（digital humanities）这个新领域的历程是这样开始的——2007年，我尝试探讨“中国历代人物传记资料库”（CBDB）的资料和数据结构可以为研究南宋（尤其是12世纪）的书信网络带来什么样的资源。我在那一年的“中古中国的群体传记学：对中国历代传记资料库的利用”（Prosopography of Middle Period China：Using The Chinese Biographical Database）工作坊<sup>②</sup>提交了一份报告，认定CBDB不适合用来研究书信网络，因为它

① 参见：<http://chinese-empires.eu/blog/digital-interpretations/>。

② 工作坊的内容可参见：<http://www.songyuan.org/Warwick.htm>。

的相关资料还远远不够。我同时指出它应该进一步扩充，收录更多现成的工具书，并容纳更个性化的小规模资料。在这种探索和研读数据库开发工作的相关文献的过程中，我越发意识到把我本人搜集的文献和阅读笔记跟CBDB结合，能起到很大的作用。

我在这方面的尝试，源于我对宋人笔记的编纂和刊刻的社会文化史研究。这是我对宋朝时政和历史类文献（例如邸报、会要、地图、奏议等）的接受史进行考察的一部分。我首先编了一个表格，录入目前所见宋代刊刻笔记的所有版本信息。<sup>①</sup> 我把这个表格变成一份统计表，其中栏目包括作者、笔记原题、编纂时间、刊刻时间、刊刻者、赞助者（如果有的话）、刊刻地点、目前收藏地点（如果该笔记流传至今）和这条信息的出处。有了这些信息，再把它们连到CBDB不断增加的人物传记资料，我得以对很多宋人笔记的作者及其生平进行社会背景和地理的分析。我也可以借助哈佛大学地理分析中心的贝明远（Lex Berman）及其同事管理的“中国历史地理信息系统”（CHGIS），运用其提供的地理坐标来画地图，呈现多部宋人笔记的序文和各书目材料中提到的地名。根据宋代刊刻笔记的社会情况，我在一篇论文中指出，从这些数据中，我们可以看到高层官员对笔记的编写和刊刻仍有重要作用，但他们扮演的角色不再像过去那样占据主导地位。<sup>②</sup> 而这些笔记的主要内容却没变，仍是关于朝廷政治。因此，我所看到的情况和张晖等学者的观察不同——他们认为笔记的作者逐渐变得不再是高官，其中包括了更多地方士人，其中有的人甚至没有官位。<sup>③</sup>

在把自己搜集的数据跟其他数据库连接起来之后，我继续扩展这种研

<sup>①</sup> 网址为：<http://chinese-empires.eu/static/media/uploads/database/index.html>。

<sup>②</sup> 参看我的“The Production and Circulation of ‘Written Notes’ (biji),” in Michela Bussotti and Jean-Pierre Drège eds., *Imprimer autrement: Le livre non commercial dans la Chine impériale*, Geneva: Librairie Droz, 2015, pp. 19–47. 全文见：[https://www.academia.edu/1640813/2015\\_book\\_chapter\\_The\\_Production\\_and\\_Circulation\\_of\\_Written\\_Notes\\_biji\\_.In\\_Imprimer\\_autrement\\_Le\\_livre\\_non\\_commercial\\_dans\\_la\\_Chine\\_imp% C3%A9riale.\\_Ed.\\_Michela\\_Bussotti\\_and\\_Jean-Pierre\\_Dr% C3%A8ge\\_.pp.\\_19-47.\\_Ecole\\_pratique\\_des\\_hautes\\_% C3%A9tudes.\\_Geneva\\_Librairie\\_Droz\\_2015](https://www.academia.edu/1640813/2015_book_chapter_The_Production_and_Circulation_of_Written_Notes_biji_.In_Imprimer_autrement_Le_livre_non_commercial_dans_la_Chine_imp% C3%A9riale._Ed._Michela_Bussotti_and_Jean-Pierre_Dr% C3%A8ge_.pp._19-47._Ecole_pratique_des_hautes_% C3%A9tudes._Geneva_Librairie_Droz_2015)。同时可以参看我的 *Information, Territory, and Networks: The Crisis and Maintenance of Empire in Song China*, Cambridge, MA: Harvard University Asia Center, 2016, 第六章。

<sup>③</sup> 此处提到的是张晖《宋代笔记研究》，华中师范大学出版社，1993。——译者注

究方法，开始处理笔记的电子全文。把数据连接到这些全文，可以让我在研读笔记的过程中看到更丰富的信息。我也因而发现，要分析王明清《挥麈录》的信息网络，最好是利用笔记的文本建一个数据库。王明清经常在这部书里提到他跟谁有过对话、他的消息来自何人、他读到的书作者是谁，还有他拜访了哪些藏书家，等等。我对全书的电子版做了标注，对书中的每条消息来自何人，关于什么话题，或关于什么文献，都一一做记录。当我这样做的时候，我还在标注里为笔记提到的每个人加上 CBDB 的人名编号，然后把电子文本连到 CBDB。这样就可以从 CBDB 导出上百条在《挥麈录》提到的人物的年代、籍贯、仕途等信息。尽管我有时候要为其增补一些额外信息，或纠正其中的一些资料（类似做法往往是必要的），但 CBDB 的好处在于，它收录了昌彼得、王德毅等编的《宋人传记资料索引》<sup>①</sup> 和李之亮编的《宋代京朝官通考》<sup>②</sup> 等参考资料，而这些资料可以帮助我们重现王明清的信息网络，并分析其社会、地理面向。<sup>③</sup>

这项研究是我的新书《信息、疆域、网络：宋帝国的危机和维系》(*Information, Territory, and Networks: The Crisis and Maintenance of Empire in Song China*) 第二部分包含的内容。我在书中提出，中国政治体生成并散布信息的方式在宋代发生了结构性的转变。本来，各种体裁的文献，如疆域图、会要、邸报、军事地理书等都由中央官员编纂，并由同一批人使用、参考。不过，到了 12 世纪，阅读和评论这些文献的人是来自宋帝国在各处居住的各种文化精英。我在制度变化、法律史、文化生产等方面分析朝堂和士人之间在政治沟通上的变化。我系统地分析了笔记和其他文献中记载士人阅读各种材料的做法，并在此基础上提出士人生成文本数量的大规模上升显示两种特点，对其后的中国历史在大规模帝国的形成和维系有重要作用。首先，士人在他们对中国整体的讨论中普遍表达出一种帝国观，一种对理想中中华共同体的愿望。其次，这些文本在规模颇大的圈子中流传和被人讨论，涉及的人数众多，背景多元，分布广。正是利用了数

<sup>①</sup> 昌彼得、王德毅等编《宋人传记资料索引》，鼎文书局，1974 ~ 1976。

<sup>②</sup> 李之亮编《宋代京朝官通考》，巴蜀书社，2003。

<sup>③</sup> CBDB 的资料来源可以参看：<http://projects.iq.harvard.edu/cbdb/cbdb-sources>。这些参考资料当然不会全无缺漏，但它们都是相对重要的。

字化手段，使我得以看到官方文书的接受和士人网络呈现何种具体情况。

然后我进一步拓展这种尝试，和徐力恒、陈韵如、李允中、尤紫锡等研究生选读了更多笔记。我们的原始数据可以在我为这部书架设的网站上下载。<sup>①</sup> 在何浩洋博士的协助下，我也把这些文本和数据建成一个在线平台，为笔记中的信息网络提供互动的可视化功能（见图1）。读者可以据此验证我书中的论点，也可以对数据和文本进行其他分析，比如可以：

- 画出热点图（heat maps）和群集图（cluster maps）呈现信息来源人的籍贯和为官地点等
- 比较不同笔记提到的人的时代分布
- 查阅信息来源人是否有为官经历，以分析他们的背景
- 查阅不同信息来源人和其他人物在笔记中出现的次数
- 查阅这些人在哪些笔记段落中出现，并查阅原文

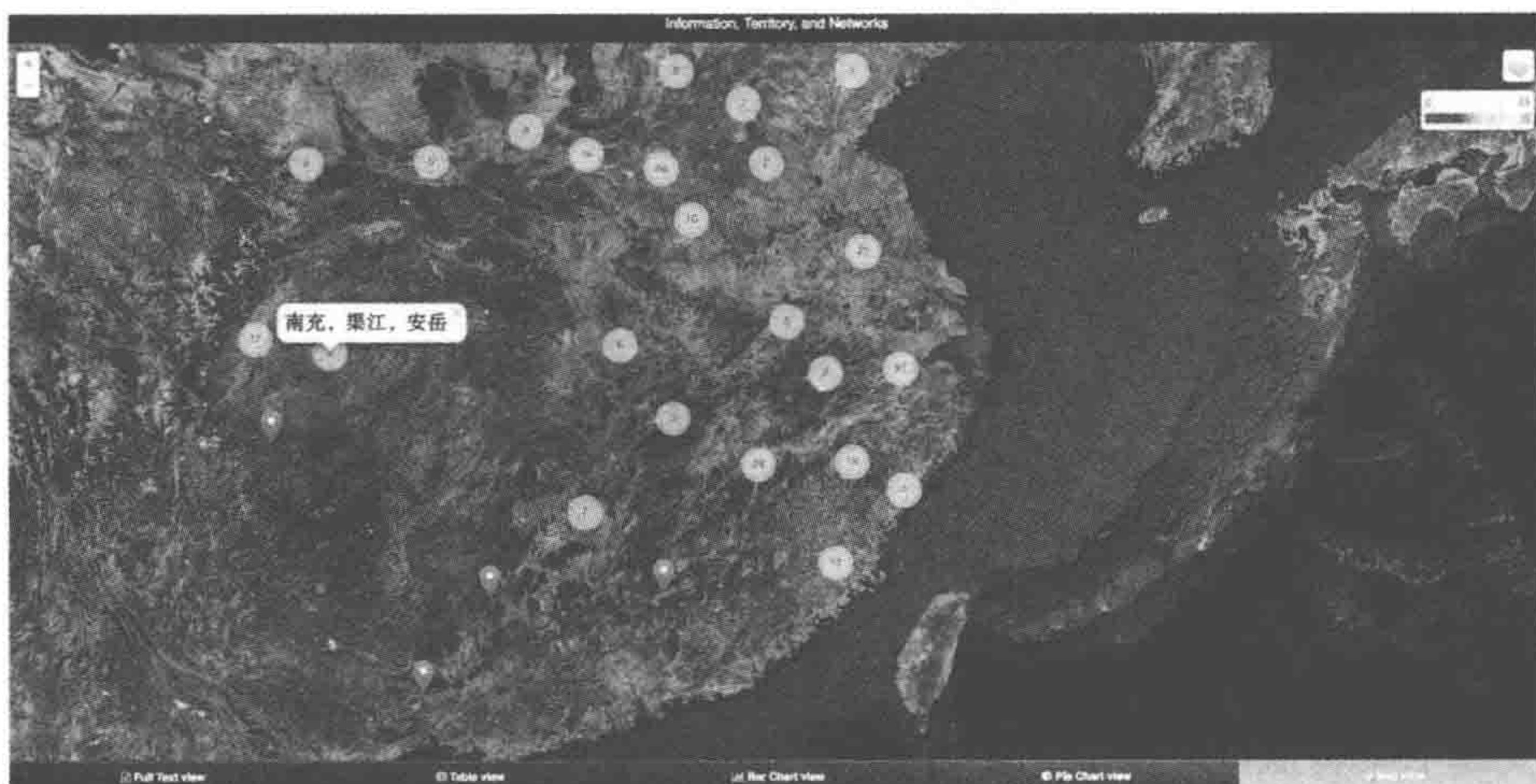


图1 为特定宋人笔记建立的互动分析平台

当前界面为“地图视图”，显示信息来源人的籍贯。最下面的工具栏除“地图视图”外，包括“全文视图”、“表格视图”、“柱形图”和“饼图”等选项。

在探讨笔记中提到的社会关系时，我对笔记这一类文献和它在宋代的变化有了新的认识。这种语境化的阅读方式，连带其他种类的数字化阅读

<sup>①</sup> 网址为：<http://chinese-empires.eu/reference/information-territory-and-networks/>。

方式（例如我书中第 8 章对不同版本进行比对的语料库语言学分析 [corpus linguistic analysis]），改变了我原来细读这些文献时得到的认识。这些方法使我能够重新检讨王明清等人在笔记中如何表达他对当时宋金关系的独特见解，并将之放到当时士人普遍看法的脉络中。学者进行这种脉络化的工作，总是很有挑战性的。

## 二 大规模的文本挖掘

我最初探索数字化研究方法时，以大规模和中等规模的史料挖掘的做法为主。我当时集中研究了特定文本（例如一位作者编纂的一系列笔记）或少数几个作者的特定著作。不过，数字化研究方法其实可以让我们做更多元化的工作。对我来说，数字化阅读能为我们实现一些独特功能——它允许学者的眼光在不同规模之间切换，可以时大时小。其实，学者已经习惯对艺术品和文物采取这种方法，他们有时放眼在其中的细部，有时同时研究一大批物件。<sup>①</sup> 而我们对文字材料也可以做同样的事情。数据审编 (digital curation, 又译数据管理) 的做法让我们得以扩大或缩小研究的规模，不过我们目前还没找到最妥善地处理数据审编的方式。

对历史学者来讲，只要想到涉及一个人群、一个广阔的地域，或各个世纪的长时段，规模的重要性很明显。我会从目前进行中的研究计划中举出两个例子，解释数字化研究方法如何让我们进行大规模的研究。第一个例子和政治史有关，<sup>②</sup> 第二个则和城市史和科技史相关。

宋代政治史和思想史的研究者长期以来热烈讨论 11 ~ 13 世纪的大规模党争。对这个关键问题的大部分研究著作要么集中分析编年史料对这些事件的刻画，要么集中在一些人物之上（有时候是重要人物，也有时候是比较边缘的人物，例如黄宽重先生近来重点研究的孙应时）。利用这些角度研究党争问题，不容易理解集体人群是怎么聚合起来变成政治联盟的。每

<sup>①</sup> 例如 Florian Kräutli 开发的工具，网址为：<http://www.kraeutli.com/index.php/2016/04/08/timeline-tools/>。

<sup>②</sup> 参看我在莱顿大学的就职演讲，全文见：<http://media.leidenuniv.nl/legacy/oratie-de-weerd-eng.pdf>。中文翻译见魏希德《重塑中国政治史》，《汉学研究通讯》2015 年第 2 期，第 1 ~ 9 页。

个士人在生涯中的不同阶段都参与到人脉的建立和维系。不论是准备科举考试和应考的时候需要这样做，寻求官职和迁转，甚至是投身其他事业，都需要这样做。既然这种活动对于士人的仕途来讲是必需的，而且是种普遍现象，那么历史学者自然不只需要了解中央朝廷中的党争，也需要知道地方上的情况如何。我认为，要研究党争如何渗透到地方上去，必须将现存的相关史料一网打尽。因此，在一些博士后和博士生的协助下，我开始分析不同党籍人士的相关著作中同时提到过哪些人名，看看这能如何帮助我们研究这个问题。通过分析 12 世纪的三个关键时刻，我们希望探讨党争在宋代的社会和文化演变中发生了什么变迁这个问题。

在一次实验中，我们整理了《元祐党籍碑》（12 世纪前十年）、秦桧（1090 ~ 1155）迫害的人员名单（12 世纪 40 年代）和庆元党禁（12 世纪 90 年代）等三个名单中的所有人物，然后搜罗每个时代在世士人的所有传世文章。<sup>①</sup> 然后，我们分析三个名单中的人在这些文章中有哪些是同时出现的。由于坊间已经有大量材料被电子化，所以这种大规模的分析是可以做到的。我们以 CBDB 的“指教年”（index years）为标准，<sup>②</sup> 囊括了 1104 年、1142 年和 1196 年这三年前后各三十年的所有宋人文集中的篇章。1104 年前后三十年的相关史料包含 59969 篇作品，包括 2231 位作者的 23701759 字；1142 年前后三十年的相关史料包含 47040 篇作品，包括 1139 位作者的 18780575 字；1196 年前后三十年的相关史料包含 52593 篇作品，包括 2598 位作者的 23446605 字。由于电脑对文本进行自动检测的技术仍有很大局限，这些数据需要经过人工仔细处理，拿掉不相关的资料。

我们据此可以得出三组文集史料都出现的人物究竟有哪些。我们得出这些数据之后，可以用来做什么？其中列出什么作品论及不同党派的成员，并注明次数是多少。我们能够以此发现哪些作品是我们过去研究中没有足够重视的，应该得到更多注意。通过这些党争成员同时出现的关系，

<sup>①</sup> Hilde De Weerdt, Chu Ming - Kin and Ho Hou - Ieong, “Chinese Empires in Comparative Perspective: A Digital Approach,” *Verge: Studies in Global Asia*, Vol. 2, No. 2 (Fall 2016), pp. 58 – 69。全文见：<http://www.jstor.org/stable/10.5749/vergstudglobasia.2.2.0058>。

<sup>②</sup> “指教年”是 CBDB 推算历史人物年代的指标，标准参见 [http://projects.iq.harvard.edu/files/cbdb/files/131009\\_rules\\_for\\_index\\_years.xls?m=1438184906](http://projects.iq.harvard.edu/files/cbdb/files/131009_rules_for_index_years.xls?m=1438184906)。——译者注

我们还可以进行社会网络分析，考察党争成员的网络重合度为何，有哪些小群体，并且其中有哪些人物占据重要地位，能起到把不同人群连起来的作用。我们可以由此回到一手史料和二手材料，进一步研究这些人物和我们过去忽略的一些小群体。我们也可以进一步考察有哪些因素影响这种同时出现的人，及其中小群体的成员构成，例如籍贯、家庭关系、仕途等。另外，我们希望以后能记录史料如何呈现不同时期的党争成员之间的社会关系，并有什么样的变化。这项研究仍在进行之中，但初步结果显示几个群体在党派结构上相当不同——元祐党人名单是小群体的松散结合，庆元党禁则是紧密地和一个核心人群连在一起的，而秦桧迫害的人员之间的关系则不明显。这显示社会网络分析可以用来展示社会关系有什么影响，也可以用来反映社会关系的缺席有哪些影响。我们也打算借助中国历史的大型群体传记数据库和文本数据库，推进目前分析历史上社会网络的做法，尤其是发展取样（sampling）的方法，用以比较社会网络中共同出现的人（co-occurrence）和同时代人群的随机抽样人群。这项工作显示，道学群体是一个联系很紧密的群体，因为当我们利用同样方法抽样分析庆元党人中的一百个人在 52593 篇作品中共同出现的程度的时候，发现其要远远低于道学群体。

这些提取数据的方式对研究很多历史问题都有帮助。我和我的一位博士生熊慧岚目前正在研究修筑城墙的长时段历史。我们利用一系列“正则表达式”（regular expressions）<sup>①</sup>，描述地方志和文集中城记、门记等史料一般提到修筑城墙的文字表述，藉此提取数据。据此我们建了一套数据集，涵盖从宋代到清代修筑城墙的活动记载。这可以让我们在空间和时间上呈现城墙的修筑活动、维护、坚固程度、工程材料、使用人力、涉及、规模、地点、作用等方面。这种研究方法对其他历史学者也有用处，例如城市史、军事史专家在研究城市规划和军事科技时就可以利用这样的数据集，考察修筑城墙和火药技术发展的关系。<sup>②</sup>

<sup>①</sup> 正则表达式可以帮助我们从全文数据中提取具有特殊写法的信息。比如中文史料中常见的日期多以年号开始，后接年、月、日，学者编写正则表达式时就可以按此格式，把一批文本中的日期信息全部抽取出来。——译者注

<sup>②</sup> 例如 Tonio Andrade, *The Gunpowder Age: China, Military Innovation, and the Rise of the West, 900 – 1900*, Princeton: Princeton University Press, 2016.

这样的大型数字化研究计划可以用于大规模的考察，也可以让用户集中考察具体文本中的史料、人物、地点和修筑事件。当然，为了达到这种目标，历史学者需要面对一些新的挑战。他们要投入大量时间在数据审编之上。所有数字化计划都是如此，因为自动化方法不是完美的。另外，历史学者进行这种大规模分析的时候，可能要调整自己的期待值。我们进行数字化的研究时，需要调整研究的习惯。例如，当进行大规模分析时，可能要忍受试验和失败，也可能得容忍一定程度的不准确数据和凌乱。而且，这种工作也不可能是一人之力可以完成的。要得到新的认识，就必须得到不同专长的学者合作，一起发展适合人文研究的数字化研究方法和平台。<sup>①</sup>

### 三 关于成为开发者

也是因为体会到这一点，引发我主持开发“码库思”（MARKUS），一个用于古代汉语文本分析和阅读的平台。<sup>②</sup> 我过去在申请一个研究政治沟通的项目时，决定聘用一位计算机专业的博士后，以我们对宋代私人著述的标记为基础，主力开发可视化和分析信息网络的工具。<sup>③</sup> 当这个项目行将结束之时，我可以断定花这么多时间和资源在数字化研究之上是让我非常满意的决定。历史学者通常会把建立网站、数据库和电子平台的工作假手于他人，承包给学校内部或学校以外的开发商。这些开发商通常只能投入有限的时间，而且会利用现成模板来解决他们客户需要处理的问题。当我和一位全职的博士后、对人文研究感兴趣的计算机专家合作时，我终于可以开始处理我研究《挥麈录》时遇到的问题。这些问题属于中国数字化研究中的一些关键的结构性问题。

第一，正如我在别处指出的那样，<sup>④</sup> 20世纪90年代初以来就有一些大型商业公司建立数据库，为历史学者提供中文古籍的全文数据库。不过，

<sup>①</sup> 关于合作方面的问题，参看我在这篇博客文章的讨论：<https://cpianalysis.org/2016/06/09/collaborative-innovation-and-the-chinese-digital-humanities/>。

<sup>②</sup> 网址为：<http://dh.chinese-empires.eu/beta/>。

<sup>③</sup> 项目网站的网址为：<http://chinese-empires.eu/>。

<sup>④</sup> 我曾在以下演讲更详尽地讨论这一点。参见：<http://leccap.ingen.umich.edu/leccap/viewer/r/az07QY>。