

2021年6月11日下午，在中国人民大学数字人文研究中心和中国人民大学科研处联合主办的第3期“数字人文无界谈”学术沙龙上，中国人民大学历史学院杜宣莹讲师和胡恒副教授围绕“中英数字史学”进行了精彩讨论。

杜宣莹：中国人民大学历史学院讲师、数字人文研究中心研究员。台湾大学历史学学士，英国伯明翰大学宗教改革与近代早期研究硕士，英国约克大学历史学博士。研究领域为英国近代史、英国国家档案与古文献学。

胡恒：中国人民大学历史学院清史研究所副教授，兼任清史研究所副所长、中国人民大学数字人文研究中心副主任、《清史研究》副主编等。2020年斯坦福大学访问学者，主持数字清史实验室及清史数据共享平台建设。研究领域为中国历史地理学、清史、数字人文等。

引言

胡恒：今年2月，杜宣莹老师以题为“门客、政治和伊丽莎白一世政权：数字人文与新视野”的研究荣获“Gale亚太学者-数字人文牛津大学奖学金”的资助，这是一个很重要的数字人文奖项。首先请杜老师介绍一下自己的数字史学研究与实践。

杜宣莹：我在英国攻读硕士与博士时，两位导师都曾参与英国国家档案数据库(State Papers Online)的建设，引导我开始了解英国档案数据库的发展沿革。近年来，我的研究聚焦于英格兰都铎晚期政权文化中的门客体制，了解到人大清史所与香港科技大学正在合作研究《缙绅录》数据库，也参加了他们组织的一些学术工作坊，于是开始关注数字人文，发现大陆的数字人文已经从传统的数位典藏向量化数据库和可视化的方向发展。而英国的数字人文立足于数位典藏，即将原始档案进行高清晰度的扫描，以为永续保存和研究便利。在利用这些档案开展研究时面临的困境是：如何在传统档案的基础上有所创新，如何处理海量的史料数据，如何让传统史料更有效率地呈现等，通过数字人文将有助于实现“档案的再现”。

对话

1.中英数字典藏各自历程

胡恒： 史学研究最重要的是获取与处理史料，将史料数字化应该是数字人文研究的第一步，这也是一般史学研究工作者最关心的。英国是如何开展历史文献数字化工作的？

杜宣莹： 数字人文在史学领域之所以引发争议，其中的一个原因在于数据可信度。数据的质与量在一定程度上影响研究的可信度，导致部分学者对数字人文的产出抱持怀疑态度。因此，若想解决数字人文研究的可信度问题，必须回归数据的根本：档案。

英国的政府档案收集始于中世纪后期，但官方档案库（State Paper Office）在16世纪晚期才建立，要求政府官员在离职或退休时将政务文书归缴入政府保存。17世纪中期，古物学者开始整理出版这些档案。19世纪，SPO改为公共文件局（Public Record Office，即现今英国国家档案馆The National Archives的前身），档案整理工作进入高峰期，政府和学界开始整理出版累积几世纪的国家档案。20世纪后期，在档案整理出版的前期基础上，史学界开始进行数位典藏，将19世纪官方的档案出版物以及16—18世纪出版的英文书册进行扫描。21世纪是英国数位典藏飞跃的开始，英国国家档案馆、大英图书馆在Gale及ProQuest等商业数据公司的帮助下，将大量珍贵的原始手稿及报纸（例如《泰晤士报》等）数字化。

英国开展数位典藏的主因有三，一是节省存储空间；二是便于研究者直接从数据库获取档案，避免档案在使用过程中的毁损；三是作为珍贵档案的永续保护和抢救工程，如大英图书馆的“濒危档案计划”（The Endangered Archives Programme）。英国在数字人文的第一个阶段（数字化扫描）做得非常好，但**如何提取原始手稿中的数据是英国数字人文面临的问题，主要有两点原因**。一是技术，中世纪和近代早期原始手稿的书写和拼写与当代语言不同，通过技术手段将其提取出来存在一定困难。二是可信度，英国实证史学的传统强调回归原始档案的判读分析。即便从原始手稿中提取出数据，英国史家对于这些数据的质疑与使用意愿也是个问题。**对原始档案的考证建构了英国实证史学的悠久传统，或许，也成为英国实现从数位典藏到量化数据库的障碍之一。**

胡恒： 中国数字典藏取得的成绩非常突出。20世纪90年代中期，台湾中研院和台湾大学就开始做数字典藏工作，即对原始的史籍进行扫描、识读，使其成为可检索的文本。**中国大陆虽起步稍晚，但发展迅速，标志性事件是1999年由上海世纪出版集团、上海人民出版社、香港迪志文化出版有限公司合作推出的《文渊阁四库全书》**。新世纪以来，大陆的数字典藏发展迅速，除了图书与文物馆藏机构的努力外，一个显著特点是商业公司的介入。如爱如生等公司，均有一些规模很大的史学数据库项目。今天从事史学研究，几乎已离不开这些数据库的帮助，由此也对史学生态产生很大影响，有利也有弊。

2.数字典藏可持续发展:商业化? 公益化?

胡恒: 看起来英国的数字典藏虽然引入了一些商业公司,但整体上还是公益性为主。近年来中国大陆数字典藏的公益性虽然也在增强,包括国家图书馆、中国社科院近代史所都有非常好的数字典藏项目,但总体而言,商业运营的数据库还是占了绝大多数。公益化或商业化的道路各有利弊,**公益化受到学界的欢迎,但如何可持续性发展尚需探索;商业化有资金、技术和人才保障,效率也很高,但归根到底要符合商业逻辑,盈利是重要目的之一,与学术研究的公益性多少存在一些冲突。**在数字典藏的发展上,英国史学界希望更多数据库公司的介入,还是由公共机构依靠财政资金的投入来开展数字典藏?

杜宣莹: 英国数位典藏的资金来源主要分为两种。第一种情况是,最初由英国政府的艺术与人文研究理事会(简称AHRC)的项目支持,但在AHRC资助结束之后,如果想持续运营,可能需要寻求商业公司的合作。第二种情况是,英国国家档案馆和大英图书馆等公共机构一开始依靠政府财政支持,但最后可能也会与大型数据库商业公司合作,寻求技术和资金支持。此外,还有第三种运营形态,即高等院校提供研究需要的档案史料,同时寻求技术合作方如英国谢菲尔德大学数字人文中心,由这样的合作方提供有偿的技术支持,且帮助资源提供方申请其他研究项目的资金,形成多方合作的关系。

胡恒: **中国大陆数字典藏发展迅速,但也存在很多不足,特别是重复建设、开放性不足、商业化过度等方面。**一些学者盼望由国家或公立机构出面进行整体规划,统一建设,但所需费用至少也要几十亿之多,短期内恐怕还很难实现。在史学领域,适度商业化仍然是当下的重要选择之一,与公益化可并行不悖。中国大陆数字典藏的公益化还遇到一个问题是不少公立机构对馆藏文献分享的意愿不足,服务和技术方面也存在较大改进空间,特别是与国外一流图书馆相比。我的一个疑问是,英国采用了怎样的机制去鼓励图书馆或档案馆将馆藏资源共享?

杜宣莹: 英国的档案馆主要有四类:第一类是英国的国家档案馆和地方档案馆;第二类是大学图书馆,保存大量地方私人或校友档案,成为研究当地文化史和社会史的重镇;第三类是教会档案馆;第四类是贵族家族文献。这些档案馆基本都是免费向公众开放,只有中世纪手抄本或原始手稿以及贵族世家文献,需要特殊程序申请查阅。目前贵族家族档案也开始通过商业数据库公司进行数字典藏。地方档案馆或教会档案馆的馆藏数字化率没有英国国家档案馆等机构那么高,所以研究地方史和文化史仍需要亲自到当地档案馆查询史料。

3.从数字典藏走向数字人文困难何在

胡恒: 杜老师提到困扰英国的一个问题是如何从数字典藏走向数字人文研究。中国其实也有同样的问题。数位典藏就史料利用而言,提供了相当大的便利。但资方,无论是政府还是科研机构,都希望馆

藏资源的数字化能够推动学术研究的革新，特别是在新问题的开拓与新方法的应用上。一旦资方觉得数字典藏并未达到预料中的成效，继续投入数位典藏或数据库建设就会有所迟疑。

人大清史所建立了**数字清史实验室**，**中心工作不是做数位典藏，而是以量化数据库和数字人文方法实践为中心**，其原因主要有五点：第一，我们自己没有多少独家的原始文献收藏，如果仅仅是将其他单位已数字化了的古籍进行再数字化，意义不是很大；第二，数字典藏对于资金和技术的需求较高，一般的科研机构没有相应的能力；第三，数字典藏大体而言本质仍然是传统史学研究方法的数字化进阶，而不是方法论意义上的转变，史学方法的革新应该更多考虑到计量和可视化等手段；第四，量化数据库的建设与计量、统计思维正是以定性研究为主的传统史学研究所亟需补充的方法之一，本身也是我们研究所需；第五，数字人文的学术评价机制是当下正在热议的问题，建设量化数据库，进而进行论文发表等科学研究，也是在现有学术评价体制内部获取认可的一条重要途径。

当下史学界有不少同仁对数字人文方法存在质疑，原因很复杂。或许原因之一是误读，一些学者仍然将数字人文方法简单地视为在数据库中搜索和理解史料的过程，进而以数据库中所收录的文献版本质量不如传统纸质文献，关键词检索方法造成史料遗漏等问题而予以批评。这种批评不能说没有道理，不过这些问题在传统史学研究中同样存在，而且它的真正原因不在于数据库而在于研究者的功力，更重要的问题也许是这种批评大大限缩了数字人文的真正内涵。数字人文不仅仅是数据库检索，而是包含了计量、可视化、社会网络分析等层次很丰富的研究方法，这些方法在其他社会科学领域长期应用并得到认可的，不应该放到史学领域就一定“水土不服”。不少对数字人文的批评实际上是误读了数字人文本身。我想问的是英国的史学研究者是如何理解数字人文的？也有学者将数字人文理解为在数据库中检索史料吗？

杜宣莹：的确，目前部分英国史学者对数字人文的理解仍局限于利用数据库检索和获取史料，数字人文仅是提供研究的便利性，而非研究方法，乃至研究思维的革新。然而，21世纪以来英国数据库的建设逐渐从数字典藏转向专题量化数据库，再加上数字人文项目的研究成果，显示英国学界已然开始产生档案思维的范式转变，即从平面的史料考证，到立体化和抽象化的数据建构。当前英国史学相关的数字人文数据库可能包含三个层次。第一个层次是论文式的人物传记数据库，如英联邦人物传记（ODNB）。第二个层次是将多种的数位典藏融合成一个综合性、主题性的入口数据库，便于研究检索。第三个层次类似于《缙绅录》数据库，如《1540-1835年英格兰教士数据库》（The Clergy of the Church of England Database, 1540-1835, <https://theclergydatabase.org.uk/>）。此量化数据库在1999年建立，最初由AHRC资助，目前累计约15.5万名英国神职人员的档案，档案来自50多个档案馆，共有约1亿5000万份资料。这个数据库更设有互动式blog，用以公布最新的数据更新动态，也以此向社会大众征求数据补充或搜集的志愿性支持。另外，这一数据库也发展出同行评议的在线期刊，鼓励使用该数据库得出的研究成果进行投稿。



图1 1540-1835年英格兰教士数据库

胡恒：看起来英国数字典藏目前主要的目的是史料获取，但也在发生一些转变。近几十年来，大陆史学界总体而言在坚守史学传统的同时，也在积极吸收国外特别是社会科学的理论和方法，许多新的方法都有试验甚至试错的空间。这可能为大陆史学界从数字典藏直接跨越至数字人文研究，甚至成为当前研究热潮之一提供了契机。

当然，新方法的起步阶段一定会带来一些争议，除了新方法本身还不成熟之外，可能还涉及思维方式的转变。传统史学对于史料的精确性追求近乎“洁癖”，一系列处理史料的传统技艺是史学研究的传家宝。只是数字人文特别是量化研究尽管对于史料的准确性也有要求，但大概还无法做到像传统史学一样，毕竟要处理的不是一种史料而是大规模的史料集，所以很多量化研究会倾向于用大规模数据来弥补个体数据的误差，用整体的统计学和概率学来研究历史的整体发展趋势。这是完全不同的一种思维，接受需要一个相当长的过程，而且对于各个断代而言，其适用范围和接受程度也存在很大差异。

4. 如何跨学科合作？

杜宣莹：人大清史所和香港科技大学正在合作研究《缙绅录》数据库，也在建设相关数据库。在数据库建设与研究中，如何克服资金和技术问题？在人文和技术跨学科合作过程中遇到了什么问题？是如何磨合的？

胡恒：数字人文时代的研究模式的确和传统史学研究存在很大差异，首先是团队模式取代了个体户模式，二是团队知识结构的异质性比较高。当越来越多的学科介入史学研究时，在方法论层面，大家可能没有统一的标准和共识，同一问题，历史学家、经济学家、社会学家共同处理就会面临方法论上的矛盾和冲突。我相信，今后的数字人文绝大多数都是跨学科的研究队伍，可能要注意的有几点。第一，未来跨学科的研究队伍将是一种常态，我们要学会如何和具有不同思维方式的人沟通；第二，要保持研究方法的开放性，既然另一个学科长期存在，其研究方法必有可取之处，所以要学着尝试去理

解和接受；第三，在跨学科的研究团队中，要对其他学科的基础方法论持尊重态度。我们既要创新，又要坚守，中间是痛苦的磨合过程。

5.史学研究中数字人文的限度

杜宣莹：我认同胡老师的最后一点，要尊重不同学科的基础方法论。数字人文是极佳的创新研究方式和思维，但当研究者进入数字人文领域时，**需要思考研究问题是否适合数字人文，而不是为了迎合数字人文而进行数字人文。**我们需要思考哪些问题能通过数字人文拓展出不同的研究视野，或基于不同目的而创新出迥异于传统的表述形式，甚至借助数字人文重塑根本的问题思考方式。需注意的是，数字人文在史学研究中仍有一定局限性。很多数据库工具的文本标注和提取并不是依照当时的历史语境进行，分析16世纪的历史文本时采用的是21世纪现代语境，却忽略这个词在16世纪的解读，包含字义、情感或使用方式，和21世纪或许截然不同，因此仍须回归原始文本进行考据分析。使用数字人文不意味可以脱离传统史学，两者并非背道而驰，而是相辅相成。

胡恒：对我们来说，从事数字人文研究，数字化是手段，最终还是要回归到传统的人文问题。数字人文可以作为一种新的方法，但目前看起来还未根本改变史学思考问题的基本方式。数字人文不是万能的，在研究方法的应用、数据的处理等方面都存在一定的局限性，不能也不会替代传统人文研究。我们只有始终明确这点，才能让数字和人文的结合更加完美。

提问：在研究中如何平衡传统史学方法论和数字人文方法？数字人文方法是不是一种工具？平时使用数据库时需要注意哪些问题？

胡恒：第一，**数字人文方法不仅是一种工具**，也是一种完全不同的思维，包括量化、关联、空间分析等，都是传统史学研究中较为薄弱的部分。第二，数字人文研究不会损害传统史学研究方法的价值，譬如就一般性的史料利用而言，经由数据库检索的资料，还要注意核对版本信息，精心对史料进行辨析，不轻信检索结果，这就需要传统史学训练。

杜宣莹：英国史研究有一些权威数据库如ODNB、SPO、BHO与EEBO等，立足于原始档案或权威史家的编撰，可信度无太大争议。其他专题数据库则必须考证档案的来源和质量以确认数据的可信度，从而决定是否能够使用。不管哪种类型的数据库，使用档案前必须进行基本信息如版本的考据，即使是权威数据库，仍可能存在错误。若通过工具平台析出数据，进行分析，也需注意档案的筛选、工具的选用、设定与可能的缺陷，例如上述的历史语境错置等问题。

6.数字人文对人文教育的影响

提问：在数字人文领域，理工科背景是否更有优势？数字人文对未来的文科教育会产生哪些影响？

胡恒：是否更有优势还不好说，但未来理工科尤其是有技术背景的人加入人文研究应是一个可能趋势。复合型人才也会更受欢迎，人大历史学院的师资队伍中就有经济学、地理学出身的老师。

史学研究的风格和方法从来不是固定的，数字时代一定也会有所变化，有的已经在发生，未来可以去想象。

7.数字人文项目的发展趋势

杜宣莹：数字人文在英国已呈现出研究、教学与公共史学的多元化发展。例如，英国约克大学Dr. John Cooper主持一个研究英国议会大厦沿革的项目“St. Stephen’s Chapel, Westminster, Visual & Political Culture, 1292-1941”，团队成员涵盖历史学、艺术史、考古学背景的学者，还有技术团队。他们通过搜集与分析文书史料、出土文物与其他档案等，3D立体复原中世纪后期迄今英国议会大厦中核心建筑St. Stephen’s Chapel的沿革。这个数字人文项目不仅产出专著、多篇期刊论文和延伸的研究项目，也和结合教学功能，培养参与项目的博士生完成学位论文。更值得注意的是，这项目也结合当前英国公共史学的趋势，建设项目网站 (<https://www.virtualststephens.org.uk/about>)，公开研究成果，尤其是互动式的3D还原影像。**数字人文项目主要有三点优势：**一是研究者可以发展跨学科与多维度的研究视角和研究产出，二是可作为教学项目，三是符合英国公共史学的发展趋势，成为公众了解历史的渠道和平台。数字人文已然冲击英国根深柢固的实证史学传统，引领档案的双重变革，储存方式从实体收藏到数字保存，档案的思维范式也逐渐从史料考证转向数据网络。这势必挑战，如同胡老师所言，英国传统史学对原始史料的“精确性”要求，但数字人文对传统档案思维的突破，必然成为问题思考方式的革新契机，拭目以待。

ST STEPHEN'S CHAPEL, WESTMINSTER
VISUAL & POLITICAL CULTURE, 1292-1941

Home About Visualization Blog Publications Contact

About the Project

As a monument to medieval kingship and a setting for parliamentary government, St Stephen's Chapel in the Palace of Westminster has helped to shape the political culture of the nation. Funded by the AHRC (2013-17), our project explores the history, art and architecture of the royal chapel which became the first dedicated House of Commons. This website provides access to the [3D visualizations](#) modelled from our research (the digital modelling was undertaken by [The Centre for the Study of Christianity & Culture](#) at the University of York).

As the king's chapel in the Palace of Westminster, St Stephen's was built and furnished over seventy years by Edward I, Edward II and Edward III, to create a setting for royal worship as lavish as any in Europe. When St Stephen's was dissolved during the Reformation of Edward VI, the upper chapel became the meeting-place of the House of Commons. It survived until 1834, when the old Palace of Westminster was ravaged by fire. In the neo-Gothic new Palace designed by Sir Charles Barry, St Stephen's Hall took the place of the former upper chapel while the lower chapel was ultimately restored as a place of worship. Decorated with wall-paintings depicting 'the building of Britain', St Stephen's remains

In this section:

- About the Project
- Project Team
- Events and Timeline
- The Building Accounts for St Stephen's Chapel, Palace of Westminster, 1292-1366
- News Archive
- Acknowledgements

Follow us

- Tweets by [VSS_Project](#)
- Follow [@VSS_Project](#)

Panorama of the Ruins of the Old Palace of Westminster, 1834, by George Scharf. © Palace of Westminster Collection (WOA 3793) <http://www.parliament.uk/art>

图2 英国议会大厦研究项目

结语

胡恒：数字人文天然具有国际性。中国史和世界史在数字资源利用、数字人文方法进展等方面既有同样的问题，也有各自的差异。未来可围绕这些问题继续交流。

[【往期回顾】DH无界谈001|哲学与计算机科学的“本体”之辨](#)

[【往期回顾】DH无界谈纪要|重访古典中国的新路径——学术立场、方法与技术工具](#)

撰文：何思源

排版：姜小艺



RUC数字人文研究中心

微信号 | rucdh201

网址 | dh.ruc.edu.cn

邮箱 | rucdh@ruc.edu.cn