数字人文研究的图书馆学方法：书目控制与文献循证*

□刘炜　林海青　夏翠娟

摘要　考察人文计算和数字人文的历史可以发现，数字人文本质上是一个“方法论共同体”(Methodological Commons)，即由人文学者采用计算机方法和工具，依靠数字化和数据化的人文资料从事人文研究的实践活动。这类实践表现出人文科学在方法论层面一种融合性和共通性：各门人文学科虽然问题不同且对象迥异，但由于材料的处理和内容的组织都采用了数字化工具和方法，因此构建统一的数字人文基础设施已成为可能。目前对方法论共同体的深入研究尚不多见，估计是由于数字人文的跨学科属性，既要熟知人文科学及其各分支学科的研究方法和一般规律，又要具备数据科学应用能力。这一领域常常是人文学者先知先觉，计算机专家随后跟进，而目前却成就了很多数据专家，成为具有知识组织能力的新型图书馆专家施展才华的领域。文章从引介数字人文的方法论基础和主要支撑技术入手，重点讨论了图书馆传统文献学中书目控制和文献循证两种方法能够对数字人文研究带来的巨大作用，这两大方法提供了数字人文的两大基础——数据与方法——的最基本的支持，并对未来图书馆在人文研究中的服务进行了展望。

关键词　数字人文　图书馆学　方法论　方法论共同体　书目控制　文献循证
分类号　G255.1
DOI　10.16603/j.issn1002－1027.2018.05.018

1　引言：人文及人文科学的产生

人文科学在东方产生于先秦（公元前221年秦统一中国之前），而西方则诞生于轴心时代（公元前800年至前200年）[①]。相传孔子“祖述尧舜，宪章文武，上律天时，下袭水土”（《中庸》），删定六经，创办私学，收门徒三千，“身通六艺者七十有二人”（《史记·孔子世家》），所谓“六艺”，即“诗、书、礼、易、乐、春秋”，皆为最早的人文科学的分类体系。而古希腊哲人与印度早期佛陀们对于宇宙、大地和人的精神的激辩与冥想，经后人整理，留下了大量著述，成为所有科学（不仅是人文科学）的肇始和基础。

人文（Humanity）一词的含义是“以人为中心，关于人的一切”。人文科学是一切有关人和人类社会的知识积累，严格说来应称之为“人文学科”，因为“科学”的严格意义通常是指自然科学，在方法的采用（如数学方法）和可验证性（如可证伪）方面有严格的规定（需符合研究范式）。文中将人文科学等同于人文学科，取其作为“研究领域”的广义解释。人文科学自诞生以来，在近2000多年的历史中，伴随人类自我意识的觉醒、将一切托付于上帝以及文艺复兴之后的再次启蒙，经历了一个产生、怀疑和再发现的过程，最终经过工业革命而进入现代文明。通常人文科学包括语言、历史、考古、文学、艺术、哲学、宗教、法律、社会学、人类学，等等，由于人文与社会科学的边界并不清晰，我们可以将所有自然科学之外的学科都归入“人文与社会科学”范畴，它们都承载着人文精神，在当今时代，对人类意识形态的影响甚至超越了宗教，成为现代人类社会的共同信仰。

* 国家社会科学基金项目“面向数字人文研究的图书馆开放数据体系构建与服务模式设计研究”（编号：18BTQ027）的研究成果之一。
通讯作者：刘炜，ORCID：0000-0003-2663-7539，邮箱：wliu@libnet.sh.cn。

① 德国哲学家卡尔·雅斯贝（Karl Jaspers）提出的著名概念，认为公元前500年前后，在北纬30度上下（25－35度之间）是人类精神的重大突破时期，结束了几千年的蒙昧时期，开创了新的以理性为主导的文明。

“人文”与“科学”等许多概念一样，从源头上属舶来品，与中国传统学术从观念到内容都表现迥异。但同为学术，其基本的认识过程和逻辑基础是相通的，最大的一致，是它们的繁盛都来自于载体化文献的大量产生。轴心时代以西方的纸草和东方的简牍为代表的书写技术都已发展成熟，积累了大量的文献。没有文献，就没有人文学说的广泛传播和学者间持续的争鸣和交流，也就很难有人文思想的传承和人文科学的进步。传统人文研究的基础是文献，探索文献源流、考订版本、编录存佚、校勘真伪、音韵训诂、辨析义理等是研究传统学术的门径和基本功，可以说这些传统的文献学方法，支撑了传统人文科学2000多年的发展。

2 本质：数字人文是方法和工具的变革

近年来，数字人文（Digital Humanities）引起了广泛关注，学界对其争议不断、褒贬不一，但其带来的颠覆性正在造成人文科学“数千年未有之大变局”。作为现代信息技术应用于传统人文研究而形成的一个跨学科研究领域，数字人文概念源自于一个非常“技术”的名词：“人文计算”（Humanity Computing），它是对人文科学各领域普遍采用新方法新工具的一种认识和归纳，这些新方法新工具有一个共同特点，就是应用了与计算机有关的各类技术，例如数据库技术和计量统计方法等，因此与其说数字人文是一门新兴学科，不如说是一种新的方法和工具体系。《中国社会科学报》也认为：大数据时代数字人文将促进方法论革新[1]，新兴的数据技术带来知识单元的细粒度化，知识组织的语义化，知识呈现的可视化，为人文研究提供了前所未有的强大工具，使当今人文研究能够见所未见、能所不能。

2001年英国布莱克威尔（Blackwell）出版公司的编辑把当时拟出版的论文集《人文计算指南》（*A Companion to Humanities Computing*）改名为《数字人文指南》（*Companion to Digital Humanities*），宣告了“数字人文”这一学术名词的正式启用[2]。2002年威拉德·麦卡蒂（Willard McCarty）和哈罗德·绍特（Harold Short）对该领域进行了深入研究，提出数字人文研究的“方法论共同体”（Methodological Commons）概念，意为“数字人文共通方法的一种抽象化和理论化”[3]，这是第一次从人文科学总体上认识到一种跨学科的、相通的数字化方法。这些早期研究引起很多讨论，反映在后来出版的《新数字人文指南》[4]和 debates in the DH [5]及其续编等论文集中。迄今学界在认为数字人文是一场通用方法论变革方面，已基本达成了共识，然而对这场变革究竟有多大影响，数字人文究竟还有多少“人文”，以及数字人文对人文科学总体上和特殊性两个方面会产生怎样的影响等，争论仍非常激烈，尚未取得统一的认识。但可以肯定的是，方法和工具的变革并不会只带来外在的、数量或规模上的变化，其对学科领域会产生从深度和广度两方面的扩展，并且会对各学科研究对象带来重新或全新的认识，目前其实已经开始颠覆旧的学科，甚至产生了新的学科。如同望远镜和显微镜的发明虽然只是新的工具，但却带来了天文物理和生物医学的大发现，许多新的事实不仅改写了过去的一些传统学说，而且诞生了大量新的理论。

还有学者从人文科学的研究行为方面找到了大量的相通之处，希望以此支持“数字人文”这样一个共通领域的存在。美国数字人文专家约翰·安斯沃斯（John Unsworth）对数字人文的通用方法总结了七个“学术原语”（Scholarly Primitives）”[6]，分别是“发现（Discovering）、注释（Annotating）、比较（Comparing）、参考（Referring）、抽样（Sampling）、说明（Illustrating）和表现（Representing）”。后来欧洲学者布兰克·托拜厄斯（Blanke Tobias）和安德森·希拉（Anderson Sheila）等将其规范为五个基本原语和多个“二级原语”，五个基本原语是“发现（Discovering）、收集（Collecting）、比较（Comparing）、传递（Delivering）和协作（Collaborating）”[7]。OCLC于2009年对科研行为的“原语”进行了比较系统的阐释，发表了 Palmer 报告[8]。“学术原语”概念的提出为数字人文的通用方法学提供了一个切入点和共同基础，所有这些“原语”研究行为，都可以通过各类软件程序和工具提供支持。

3 颠覆：数据化带来方法和工具的升级换代

数字人文是人文研究新方法新工具的集合，即“方法论共同体”，其前提必然是研究素材的数字化，正是由于人文研究赖以开展的基本材料在形式上发生了巨大变化，处理材料的工具手段发生了巨大变化，才带来研究方法上的变化和挑战。

如果我们将人文科学的研究从方法论演变角度

做一个历史考察，可以发现传统人文与数字人文的巨大不同。传统人文的各个分支领域在从古代走向现代的过程中经历了很大的变化和扩展，例如将《六经》（诗、书、礼、易、乐、春秋）作为六门学科，大致对应现代人文体系中的“语言学”“政治学”“社会学”“哲学”“艺术”和“历史学”，虽然其内涵和外延已发生了巨大变化，但当考察计算语言学、数字政治、数字哲学、计量社会学、数字艺术、计量史学等数字人文相对应的学科变化时，其变化就更加让人匪夷所思，对象可能还是那些对象，但技术方法的不同，问问题的方式和角度可能完全不同，深度和广度也得到很大拓展，最终会造成学科重心的偏移和研究疆域的扩展。

表1 传统人文的方法进化

学科	古代认知	学科目的	现代方法	数字技术	主要工具
诗/语言学	诗以正言，义之用也	研究与语言相关的现象和规律	文本（包括语音）及通过文本的交互	基于文本各种表现和行为的分析	各种文本分析（包括语音/语法/语义分析）
书/政治学	书以广听，知之术也	关于获取和运用权力的研究	思辨与逻辑分析	内容挖掘，社会网络分析	词分析、图数据库
礼/社会学	礼以明体，明者著见，故无训也；	社会与人类群体行为的理解	各类实证方法，调查研究	逻辑归纳，统计分析	调查管理，统计分析
易/哲学	易为五常之原，易不可见，则乾坤或几乎息矣，与天地为终始也	对世界的根本认识	概念化，逻辑	本体/语义技术	概念语义分析
乐/艺术	乐以和神，仁之表也	审美载体的创造和研究	多媒体	多媒体技术	各种数字媒体处理工具
春秋/历史	春秋以断事，信之符也	记载和解释人类活动进程	基于各种史料的解释和重建	内容分析	分析、模拟、统计等
新兴人文：人类学、考古学、文化分析、文化组学（Culturomics）等					

数据和方法，是数字人文的两大支柱（见图1），经过有序组织的数据并附带各种工具，即构成数字人文研究的服务支撑平台。

就像实验研究基于对实验的推理一样，人文学科的研究是建立在对资料的分析和推理基础上的，资料可能是文本的，物理的或无形的（Intangible）。传统人文的数据可以认为就是以纸张为主要载体的文献，文献可以数字化成图像或文本，布萨神父对于数字人文的开创性工作就是从文本处理和组织开始的。而实物可能对考古学、博物馆学等学科更加重要。另外还有一类容易被忽略的实体，即非物质的无形物，如情感、嗅觉、虚构人物或抽象概念等，也是人文科学（如文学、艺术、宗教等）的重要对象或素材。从研究角度来看，文本是人文数据的基本形式，是许多数字人文研究的起点，但因为它太重要，致使我们常常忽略了其他形式，这是需要避免的。数字世界是真实世界的一种映射和模拟，因此所有的数字信息都可以是数字人文的内容材料。

数字人文的数据是各种数字文件，包括各类人文研究对象的数字化文件及其元数据。物联网技术正在将人类能感知的所有信息都以数字化方式表达，在虚拟世界建立起真实世界的对应代表物（Surrogate），因此这些数字文件不一定是传统材料的数字化，越来越多的原生数字资源可直接作为当前数字人文研究的素材。

人文研究方法可以分解为研究过程、研究行为和所采用的分析技术等三个方面，例如归纳、演绎、比较等通常所说的研究方法是由研究过程和研究行为决定的，而卡片目录、逻辑符号系统、地图投影法等都是为帮助进行研究而采用的技术。数字人文在方法论方面带来的变化最大，不仅能够对研究对象进行细粒度的数字化和多维度描述，而且利用计算机强大的计算能力和永不遗忘的特性，对所涉及的实体关系进行全方位的模拟和操控，为各种研究行为（即上文中所列举的“原语”）提供工具，使整个研究过程在信息系统的掌控之下，实现自动化甚至智

慧化。

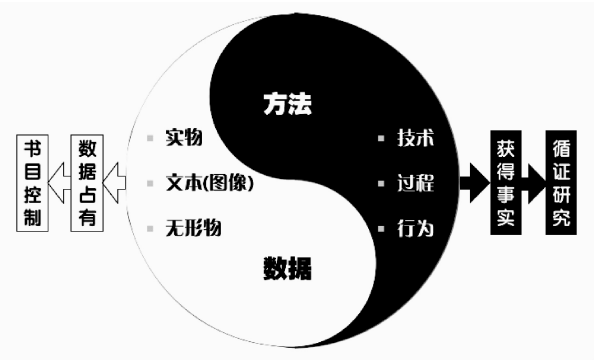

图 1　数字人文的两大组成：数据与方法

材料获取是大多数人文研究过程的第一步，能否充分占有足够的资料往往成为一项研究成败的关键。数字人文的研究素材已不是简单的文献，也不仅是经过简单数字化扫描和标引而形成的“文献数据库”，或者经过光学字符识别（OCR）或大量的人工输入而形成的海量文本，而是提取人、地、时、事、物等实体对象，描述它们的复杂属性并建立起相互之间关联关系的知识库，是真实历史和社会系统在虚拟世界中的映射和模拟，这使得数字人文能够从一定程度上“还原”现场或“重现”历史。数字人文研究往往能通过“大数据”的占有，在研究的时空范围和广度深度方面大大超过传统研究，可以进行“全数据”研究。这当然有赖于海量数据的获取、加工和处理，以及完善的数字人文基础设施的建立。数字化加工处理的方式和所建立的检索系统极大程度地决定或影响着提问的角度和深度，甚至可以说有什么样的数据就能做什么样的研究。

传统人文研究的方法相对来说较为简单，通常是掌握了事实之后，通过分析、概念化、类比、归纳、演绎、综合等，即可以得出结论，其基本要点是基于事实而符合逻辑。数字人文在研究流程上能够突破传统人文的线性模式，更易于大范围协作，能够通过不断增添事实论据，进行反复比较论证，并通过社交网络进行更充分的交流，成果形式也更加丰富多彩。数字人文从两个方面促进了人文研究方法的科学性：一是对于事实的掌控，通过“数据化”使基本事实颗粒化，能够更加方便地进行逻辑一致性检验；二是通过提供大量的计量统计分析手段，更容易定量化。近年来数据科学发展迅猛，已形成一整套数据管理方法论体系，其中本体技术、语义万维网相关技术（如关联数据）等是专门针对知识单元（语义单位）提供解决方案的技术，有学者将这类数据称为“智慧数据”[9]。这些新技术不仅带来自然科学研究向“第四范式”变革，也同样影响了人文领域走向数据密集型或数据驱动型研究。

4　方法：书目控制与文献循证

4.1　作为人文研究入门之径的文献学

把人文科学当作一个整体来研究，探寻学派源流、考证学说出处的文献学是一个无法回避的领域。传统文献学多从物理形态和版本特征方面考证，涉及内容的流传、真伪的考证、内容的比较等，几乎是所有人文研究的起点。应该说自从有了文献和学术传播之后就有了文献学，但这个概念的正式形成却是 20 世纪以后，东西方都是如此。1920 年梁启超在《清代学术概论》中首提“文献学”，三年后他在《中国近三百年学术史》中又说，“明清之交各大师，大率都重视史学——或广义的史学，即文献学”。中国文献学的奠基人之一张舜徽认为，文献学在中国古代多属于校雠学范畴，涵盖目录、版本等分支学科[10]。西方 17—19 世纪的古文献学也有类似的说法，涵盖 Philology（历史语言学）、Bibliography（书籍学）、Diplomatics（古文书学）或 Paleography（古文字学），涉及书目、校勘、版本、文书、文字、文体等多方面的研究，现在都可以翻译为文献学。20 世纪之后随着文献的形态由书籍向论文、报告等深化，又产生了 Documentation（文献学）等新名词和新领域。

人文研究之所以离不开文献学，正如梁启超认为广义史学即为文献学，各门人文学科其实都离不开史学一样，其具备很强的基础性和指南性，文献学的工作是校勘注释、阐发引申、去伪存真等，能够为各学科提供材料，同时提供前人的学问门径和方法论。

传统人文的研究成果主要以文献的方式存在并提供交流，文献中所记载的大量内容，如人物、史实、年代、名物、典制、天文、地理、历算、乐律等，涉及自然与社会、时间与空间等各个方面，从传统学术的角度来看，都属于“文献学”范畴，涉及目录、版本、校勘、辑佚、辨伪、文字、音韵、训诂、考据、义理等多方面的学问，这些学问有许多边界模糊和因人而异的地方，但可以从文献形式和文献内容两个方向来把握：目录学和版本学着重从文献形式方面厘清脉络；而校勘、辑佚、训诂、义理等则是提供内容方面考辨，

从方法学角度前者可称为“书目控制”，后者则以“文献循证”概括。从现代学术角度看，这些大都被归入“图书馆学”范畴，因此本文将由通过文献控制和引证发展而来的方法称为“图书馆学方法”。对现代人文研究来说，结合数字技术，可建立书目控制和文献循证的各类“知识库”，“反哺”传统学术，有望构建起具有中国特色的数字人文方法工具体系。

4.2 书目控制

书目控制是通过对各种目录、索引、摘要等二次文献的编制和利用，全面充分地了解和掌握某一主题领域或特定类型的文献，并掌握其分布和发展规律的一种学术活动。

书目控制是20世纪40年代美国芝加哥大学图书馆学教授谢拉等提出的概念[11]，借用了美国科学家维纳的控制论思想。1971年德国的卡尔特瓦瑟总结了世界书目控制实践，把对文献的控制推广到世界范围。1980年汉斯·H·威尔斯在《书目控制论：文献检索的理论》一文中，再一次比较系统地论述了控制论与书目控制、书目系统的调节与控制、文献鉴别等问题。书目控制在中国古代主要体现在目录的编撰和实践中，一直都有类似的提法和实践，如宋郑樵有云：“学术之苟且，由源流之不分；书籍之散亡，由编次之无纪”(《通志·总序》)，又说：“类例既分，学术自明，以其先后本末俱在”(《通志·校雠略》)。清代学者王鸣盛说：“目录之学，学中第一紧要事，必从此问途，方能得其门而入(《十七史商榷》)。章学诚也认为目录可以“即类求书、因书究学”(《校雠通义》)。

网络时代的书目控制是在充分占有某一领域的事实、数据和文献的基础上，按需求进行描述和组织，并监控其发展变化。具体方法是对于每一项有意义的知识单元都需要赋予独立的网络ID标识并以标准的方式(RDF)进行描述，所有数字对象(包括由简单数字对象构成的复杂对象)均按照领域规范(即所谓模式Schema)进行代码化，所有描述和组织规范即构成网络知识组织系统，这就是网络时代的书目控制，目前已有以知识本体和关联数据为代表的成熟的语义万维网技术提供解决方案[12]。网络时代的书目控制所控制的对象更加广泛，不仅包括二次文献，也包括原始对象的数字化替代物，即所有有意义的“数据”，但“控制方式”不一定是占有，分布式环境下掌握数字对象的线索，或取得访问权也是一种控制。

书目控制的范围、粒度、程度根据研究领域和需求的不同而有所不同，书目控制的主要目的是发展能够掌握与控制文献的各种手段，技术的进步可以带来更多、更广泛的控制，从而更加有助于开展基于数据的研究。

4.3 文献循证

循证研究即“基于证据的研究”，相对于基于信念(常常是偏见)、基于经验(常常不可靠)、基于伦理(常常不科学)而言，基于证据是指任何结论都需要从客观证据中得出，是科学研究的基本要求。当证据主要由事实、科研结论或数据组成的时候，由于这些材料基本都是文献形式，可称之为“基于文献”的循证研究，即文献循证。

循证研究较为成熟的领域是“循证医学”，已形成一套严格完整的原则、策略和过程方法。医疗决策通常取决于患者症状、医生经验和已有的研究结论，循证医学发展了一整套方法来保证决策的最优，主要解决医疗实践中的主观性带来的问题，剔除临床实践中的主观成分。

人文社会科学很早就开始应用循证方法，也已成为一个普遍趋势。当然人文研究无法完全排除主观因素，如个人信念或价值观的影响，但为了保证客观性和科学性，应将人为因素降低到最低程度，通过采用一定的方法论来给予保障。循证研究所制定的原则、流程和方法就是这样一种基础的保障。

文献循证主要借助各种形式的文献中所包含的事实，对研究的问题形成一定的证据链，在一系列因果关系中寻找可靠结论的最佳实践。传统的版本、校勘、考据、辑佚等都有大量的循证实践，尤其是辨伪，总结了各类行之有效的循证方法，虽然从现代科学的角度这些“循证”方法还不是很系统很完备，但至少“基于证据”和符合逻辑这两方面是一致的，这是我们可以利用信息技术加以提炼和发展的。

在文献循证中有相互矛盾或不完整的资讯时，以一种切实存在的、可以计量的或确定的资料作为论证依据。这就是循证。事实是循证研究最重要的基础，在建立知识库的过程中需要把大量的事实进行结构化并存储起来，目前RDF技术就支持这样一种海量陈述的集合，能够进行一致性检验和推理，一定的推理能力是构成“证据链”的基础，也是文献循证方法得以实现的基本技术。与此同时还有一类技

术非常重要，即真伪和可信度判别技术，需要有一定的方法对文献中记载的数据内容建立可信度指标，因为虽然对文献记录来说是客观记录，但这些记录的内容并非就一定是真实可靠的，可以通过互证、相关实体的可信度建立一个相对真实性的描述，尤其重要的是这个描述所确立的指标还需要是动态的，根据系统内容的增减代谢和用户的使用情况，或其他相关的指标变化，能够不断地进行重新衡量和计算。只有这样，来自于文献和已经积累的素材数据才能动态地发挥价值，循证系统才是一个不断进化的有机体。

5 案例：中文古籍联合目录及循证平台

中华古籍源远流长，保存了 2000 多年的灿烂文明，成为研究中国人文历史的主要材料和依据。盛世修典，历朝历代都希望厘清文脉、阐扬圣学，然而只有到了数字化技术高度发达的今天，才真正有可能赋予我们前所未有的能力，将往圣绝学一网打尽，并就典籍中的内容提供全面的关系描述和推理功能。上海图书馆正在开发中的“中文古籍联合目录及循证平台”（图 2）就是对上述理想的一种尝试[13]。

据统计，传至今日的中文古籍文献大约有 20 余万种，虽然在大数据时代这些数据从量上看不算很大，但要将其全部编列纲目、撮其要旨、统一揭示，还是有很多难以逾越的障碍，国家耗费了极大的人力财力，开展古籍普查、善本再造等文化工程就是国家层面的努力。在网络时代，仅仅有国家工程自顶层向下进行建设是不够的，国家的需求未必是学者的需求，必须建立一种机制，使各收藏机构、研究团体、学者用户等一方面满足自身管理古籍、提供服务的需求，同时另一方面开展协作，开放目录并共建共享，从而实现各取所需。过去这种机制只能依靠政府的号召或各类联盟机构的觉悟，而互联网时代则可以通过建立一定的共享平台，将互惠共利的机制具体化、操作化，在满足各联盟机构各自的需求的同时，方便高效地开展合作，这个平台也能在服务中得到良性发展、自我生长。

上海图书馆“中文古籍联合目录及循证系统”（以下简称“上图古籍系统”）的首要目标是建立一个大而全的目录系统，除了提供各参与馆的古籍目录查询之外，还能够提供全国古籍联合目录功能，以及曾经编撰并流传至今的重要古籍目录，即史志目录、

图 2 上海图书馆“中文古籍联合目录及循证平台”

官修目录和私家藏书目录等。大而全的目录系统是书目控制的必须。目前上海图书馆利用自己和第一批合作机构（加州伯克利大学东亚图书馆、哈佛燕京图书馆和澳门大学图书馆等）的馆藏古籍数据开发一系列数据转换、清洗、校验工具软件，新增机构可以用来快速转换数据并迅速建立自己的古籍门户。这些软件有从 MARC 或表格转换成符合本系统书目本体的 RDF 数据，各类规范名称的加工、匹配与识别工具，以及利用已有的目录数据进行古籍编目或完善记录信息的工具等。

很多图书馆已经有自己的古籍系统，数据大都是 MARC 格式，另一些尚未提供古籍查询服务的机构也有简单的古籍登记账册（如古籍普查目录），可以表格或文本方式提供。该目录系统建成后可望提供一个较为全面完整的机读联合目录，附加大量的规范控制功能（如人名、地名、官职名、朝代、事件、分类主题等），并与历史上的各类目录建立关联关系。实现了这一点，基本就实现了古籍的“书目控制”，系统基本就能从整体上描绘出中华典籍的宏观结构，同时又可以从任何一个角度深入到微观细节。至于古籍中的特殊主题和文献类型（如地方志、碑帖等）的书目控制，有赖于通过进一步提取这些文献中的各类特殊属性，或提高标引深度来实现。目前的系统架构有足够的灵活性，允许将来不断进行内容完善和实现更加复杂高级的功能。

至于循证功能，目前还在尝试和完善中。基本想法是将古籍目录中的数据尽可能细粒度化并单元化，然后外挂各类知识库系统提供关联，形成一个智慧数据服务系统。该系统支持任意字段的空检、二次检索，支持分面过滤，支持知识检索（借助概念词表）和各种聚类，支持关联关系查询，支持知识库搜

索(SPARQL)、根据本体推理以及结果的多种可视化。

知识库系统可以借鉴当前业已成熟的人工智能机器学习技术，至少对于中文古籍图像文本识别、名称实体识别和自动关联关系发现等，可以集成到平台上去。这样不仅可以进行书目或相关实体之间的关联关系查询，还能对有关内容提供可视化，或"情境敏感型"数据显示，甚至能够具备一定的推理功能，发现未知知识或找到存在逻辑矛盾的数据。平台将会开发分面检索和组合过滤功能，并集成各类文本和内容标注、操控、"遥读"(Distant Reading)、统计与分析工具、社会关系分析工具等，所获的结果都可以以历史地图或各种可视化方式动态呈现。平台直接支持人文研究的通常过程，并能实现各类"学术原语"所定义的研究行为。这样各类人文研究项目，如编制书志、别录、文集，比较版本，校勘、辨伪、考据等，都可以直接利用平台完成，甚至直接产生一定格式的研究报告，在平台上交流或投稿，并在未来的出版环境中得到讨论和引用。

循证的实现有赖于大量的事实，这些事实不能仅仅让人来检索查询，而要以一定的编码形式让机器代理(Agent)能够"理解"并应用到查询、比较、推理、论证过程中去。上图古籍系统集成了刻工、印章、避讳的相关事实，对藏家生平轶事、典籍聚散存逸以及版本流变延续都以结构化知识库的方式挂接在系统中，同时将用户也作为系统的宝贵资源，对用户进行画像和管理，在充分保护用户隐私的前提下，将用户行为和用户贡献内容进行服务优化，结合资源画像和语义描述所得到的智慧数据，实现智慧化的数字人文平台服务。

6 结语：过往未往，未来已来

当今的人文研究的主体材料还是纸媒遗产，研究者主体也还是传统学者，其中相当一部分对数字人文的兴起极度不适应，正在成为纸张文明的殉道者，而另一部分已经是数字人文的先锋。

数字人文是一种全新的人文，它以工具的创新和方法的变革为肇始，以学科的融合和内容的颠覆为结果，虽然脱胎于传统人文，还在蹒跚学步，但已势不可挡，在整个人文疆域攻城略地，逐渐站稳脚跟。

随着数字社会的到来，数字化材料将让位于原生数字资源，但人文研究的基本流程——从占有材料到分析比较到得出结论——不会改变，数字人文的研究范式——从文本化到概念化到模型化——依旧适用，伴随图书馆向数字图书馆的转型，图书馆依旧为人文研究提供基础设施，以书目控制和文献循证为特征的图书馆学方法依旧是人文研究起点和基本方法。

中文世界的数字人文由于素材的文本化不足，以及传统研究方法的特殊性，呈现出一定的"中国特色"。主要是大量的典籍不能公开获取，而且都是图片文件，尚未文本化，这就使得人文研究无法像西方那样主要采用文本和内容分析方法，文本的缺失也拖累了建立领域本体所必需的数据化和概念化的进程，使得规范控制和循证方法无法普遍采用。这种现状使我们的数字人文研究比西方要落后不少。但如果基于图像的分析和检索能够取得一定的突破，可能会呈现出与西方数字人文不同的技术方法路径。

人文学科最初是一个统一整体，古希腊和春秋战国诞生时即享有崇高地位，但到了工业时代，科学主义和重商主义盛行，人文的分野也逐渐繁复，渐渐偏离宗旨而不受重视。数字人文打破学科的藩篱，让人们重新发现所有人文是一家，相通相持，同源同宗，但是否能重拾昔日荣光，还有赖于是否能大大促进其发展，有赖于学贯古今、横跨各科的大学问家出现。古往今来，人文大家往往出现于"雅典学园"这样的图书馆机构，当今只能出现于"数字图书馆"。未来是不是这样，让我们拭目以待。

参考文献

1 耿雪. 数字人文促进研究范式变革[N]. 中国社会科学报，2016－05－25(002).
2 Schreibman S, Siemens R, Unsworth J. A Companion to Digital Humanities[M], Oxford: Blackwell, 2004. [2018－07－22]. http://www.digitalhumanities.org/companion/.
3 McCarty W, Short H. Mapping the field[C/OL]. [2017－03－20]. Paper given at an ALLC meeting in Pisa, 2002. [2018－07－22]. http://www.allc.org/content/pubs/map.html.
4 Schreibman S, Siemens R, Unsworth J. A New Companion to Digital Humanities[M]. 2nd Edition, Oxford: Blackwell, 2016.
5 Kirschenbau M. What is Digital Humanities and What It Doing in English Departments, in Debates in the Digital Humanities[M]. Minneapolis: University of Minnesota Press, 2012: 3－6.
6 Unsworth J. Scholarly primitives: what methods do humanities researchers have in common, and how might our tools reflect this? [EB/OL]. [2018－07－22]. http://people.virginia.edu/~jmu2m/Kings.5－00/primitives.html.

7 Anderson S, Blanke T, Dunn S. Methodological commons: arts and humanities e-Science fundamentals[J]. Philo-sophical Transactions, 2010, 368(1925): 3779.
8 Carole L. Palmer, Lauren C. Teffeau, Carrie M. Pirmann. Scholarly information practices in the online environment themes from the literature and implications for library service Ddvelopment[EB/OL].[2018－07－22].https://www.oclc.org/content/dam/research/publications/library/2009/2009－02.pdf.
9 曾蕾，王晓光，范炜.图档博领域的智慧数据及其在数字人文研究中的角色[J].中国图书馆学报，2018，44(1)：17－34.
10 张舜徽.中国文献学[M].上海古籍出版社，2005：1－5.
11 M. E. Egan, J. H. Shera. Prolegomena to bibliographic control [J]. Journal of Cataloging and Classification, 1949, 5(2): 17－19.
12 Cuijuan Xia, Wei Liu. Name authority control in digital humanities: building a name authority database of Shanghai Library [EB/OL].[2018－07－22].https://doi.org/10.23974/ijol.2018.vol3.1.68.
13 夏翠娟，林海青，刘炜.面向循证实践的中文古籍数据模型研究与设计[J].中国图书馆学报，2017，43(6)：16－34.

作者单位：上海图书馆，上海，200031
加州伯克利大学图书馆，伯克利，美国
收稿日期：2018 年 8 月 15 日

Bibliographic Approach to Digital Humanities: Authority Control and Evidence-Based Practices

Liu Wei　Lin Haiqing　Xia Cuijuan

Abstract: By investigating the history of Humanity Computing or Digital Humanity, we can find that digital humanity is a kind of activities consisted of so called "Methodological Commons", which uses computers and digital tools upon digitized or datalized materials, to accomplish humanities researches. This kind of practice shows a lot of integration and commonality at the methodological level. Although they usually have different problems to solve, in different subjects with different objects, but because of the procedure and organizing approaches are of the same, so a unified Digital Humanity infrastructure has become possible. At present, in-depth research on the methodological commons is still rare due to the interdisciplinary nature of digital humanities. It is necessary to be familiar with the research methods in general and at the same time with its various branches, as well as the ability to apply data science. This field is often coming with the forerunner of humanity scholars, then computer experts follow up (now it has made a lot of data experts), and finally has become a field of library and information experts with web knowledge organization ability to display their talents. This paper starts with the introduction of the methodology and supporting technologies for digital humanity research, and focuses on the effects of bibliographic control and evidence-based practice. These two methods provide the fundamental support to digital humanities. The article also looks forward to the future library services in humanities research.

Keywords: Digital Humanity; Library Science; Methodological Commons; Bibliographic Control; Evidence-based Documentation

(接第 128 页)

The Role of Sanctity in the Information Service of Libraries

Zhang Jieqi　Zhao Keran

Abstract: Under the era of information, the sanctity of libraries needs to be understood and emphasized again, which is helpful for libraries to define themselves accurately and for librarianship to get out of the difficult position. Beginning with the relationship between belief and acquisition, some philosophers' and theologians' viewpoints are introduced to demonstrate the necessity of belief in the process of seeking knowledge, the reason why libraries need sanctity and the source of it. Libraries should become a space where the sanctity dots the gap of rational thinking and provide a trust framework for the people who are running risk to seeking knowledge. Some suggestions about the sanctity construction are offered.

Keywords: Sanctity; Belief; Library