

图档博领域的智慧数据及其在数字人文研究中的角色

曾 蕾 王晓光 范 炜

摘 要 近年来,随着“大数据”的飞速发展,一个重要却鲜为人知的概念“智慧数据”应运而生,智慧数据已经并将持续在数字人文领域发挥巨大作用。图书馆、档案馆和博物馆(简称“图档博”)所拥有的数据资源是数据时代各个领域,尤其是数字人文领域的无价之宝。如果采纳大数据的模式和思维方式、智慧数据的实现方式,以非结构化数据到结构化数据的组织和整合过程为手段,产生机器可理解并可采取行动的、一源多用、高效率运作的的数据,图档博以及相关行业将携带这些丰富的资源进入数字时代的主流。本文在阐释智慧数据的概念、方法论的转变、数字人文及其与图书馆关系的基础上,通过一些范例来展示信息服务的新思路,特别是针对文本型和非文本型原始数据的结构化和语义化处理新方法,由此证明:在语义网和大数据时代,图档博机构不仅是智慧数据的提供者也是直接受益者,智慧数据建设不仅能有效促进数字人文的发展,也将成为图档博机构最重要的新兴工作。图7。表1。参考文献41。

关键词 智慧数据 大数据 数字人文 图档博数据 结构化数据

分类号 G253

Smart Data from Libraries, Archives and Museums and its Role in the Digital Humanity Researches

ZENG Marcia Lei, WANG Xiaoguang & FAN Wei

ABSTRACT

Along with the rapid development of “Big Data” in recent years, an important yet lesser-known concept “Smart Data” has also emerged. In the context of the multiple “V”s of Big Data (Volume, Velocity, Variety, Veracity, and Value), the realization of the last “V”, Value, depends on “Smart Data”, i.e., the ability to achieve big insights from trusted, contextualized, relevant, cognitive, predictive, and consumable data at any scale, great or small. This paper first explains the rationale, definition and connotation of the concept of Smart Data through the results obtained from literature review and case studies, while exploring the current approaches according to the content of the Smart Data Conference held in the U.S. in recent years. Smart Data provides value by dealing with challenges posed by the volume, velocity, variety and veracity of big data (and resulting actionable information), as well as by improving decision making. Smart Data represents the way in which different data sources (including Big Data) are brought together, correlated and contextualized, analyzed and interpreted, in order to feed decision-making and action processes. Furthermore, this paper presents the research fields, resources, and methods collected from the

通信作者:曾蕾,Email:mzeng@kent.edu,ORCID:0000-0003-0151-5156 (Correspondence should be addressed to ZENG Marcia Lei, Email:mzeng@kent.edu, ORCID:0000-0003-0151-5156)

documents of American and European digital humanities research projects in the past seven years, alongside the topics and academic domains of the contributions and presentations at the previous five international Digital Humanities conferences. It reveals how the digital humanities have embodied Smart Data and Big Data concepts and approaches, which demonstrate an emerging and significant change in terms of methodology. The evidence indicates that: Smart Data has been, and will continue to be, playing a gigantic role in the field of digital humanities; the data resources owned by libraries, archives and museums (LAMs) are invaluable in all research areas, especially the digital humanities, in the data age. Consequently, this paper assesses the relationship between the fields of digital humanities and libraries, and the relationship between digital humanity research and data resources from LAMs. The paper supplies a number of cases that reveal new ideas for information services, especially the structuralization and semantic enrichment of raw data. In addition to showing how the structured data provided by LAMs can infinitely enrich knowledge graphs and associated datasets and be used for the development of knowledge bases, the article also addresses the data exchange and in-depth semantic annotation of images, introducing the International Image Interoperability Framework (IIIF) APIs and a study of semantic-based deep image annotation. Lastly, this paper focuses on intangible cultural heritages and the potential for taking digital humanities approaches, technologies, and other channels to promote the construction of digital resources for them, in compliance with the requirements of Smart Data. Overall, the literature review, project analyses, and various case examples used in this paper provide evidence that, by taking the concept and methodology of Big Data and using the approaches of Smart Data, we can turn unstructured data into structured data in the data organization and integration process, producing the kind of data that will be machine-processable, reusable for multi-purposes, and highly efficient in processing. Thus, LAMs will be able to bring these rich resources into the mainstream of the digital age. In conclusion, in the Semantic Web and Big Data era, LAMs are not only the providers but also the direct beneficiaries of Smart Data. The development of Smart Data can effectively enable the advancement of digital humanities, while also becoming the most important emerging work of LAMs. 7 figs. 1 tab. 41 refs.

KEY WORDS

Smart data. Big data. Digital humanities. LAM data. Structured data.

0 引言

走向 2020 的年代是数据的时代,数据已成为基础性社会资源。作为社会生产力的核心要素之一,数据在数量、质量、形态、作用等多个方面正在发生翻天覆地的变化。首先,大数据浪潮带来的不仅是数据量的激增,还有显著的国家战略层面的投资和保障,以及跨国跨领域的巨大数据应用。各行各业的专家和政府官员都在努力运用大数据解决重大现实问题。第二,

语义网(Semantic Web)的快速发展和 W3C 发布的一系列技术标准持续推进数据质量在结构化和语义化方面的深入和优化。从字符串(Strings)级别的超链接到事物(Things)之间的关联,数据所能表达与揭示的内涵越来越丰富。第三,关联数据(Linked Data)技术的成熟使得机器可理解和可处理的高质量数据集发布变得日益便利,由此大大促进了结构化与语义化数据资源的建设和再利用。与之相应,数据管理技术新格局也逐渐形成,以图数据库、键值数据库、列式数据库、文档数据库为代表的 NoSQL 类

数据库,与传统的关系型数据库互为补充,满足了不同场景下数据管理与利用的多元化需求。资源描述框架 RDF (Resource Description Framework) 三元组存储 (Triplestore) 与 SPARQL 查询语言已经成为语义数据管理的技术基础。第四,在数据应用与知识服务方面,知识图谱技术正在快速普及。从基于文本的关键词匹配和传统信息检索发展到智能化的知识检索,离不开合理的领域概念建模,由此导致知识本体和元数据描述模型数量飞速增长。搜索引擎巨头借助大型本体和元数据标准 schema.org,鼓励全球的站主 (Web Master) 在网页内直接建构带有语义的结构化数据。第五,人与数据的共生机制已经成形,从资源创建角度来看,大众直接创建、分享、整合与再利用数据已经十分普遍。从文化建构角度来看,参与文化 (Participatory Culture) 已经被社会接受。大众参与活动已经进入很多领域的工作流程,用户行为数据正在被分析利用,社会化网络则是这些数据的最直接来源。

随着“大数据”的飞速发展,一个重要却鲜为人知的概念“智慧数据”应运而生。那么,什么是“智慧数据”? 图书馆、档案馆和博物馆 (以下简称“图档博”) 如何借助大数据和智慧数据并以前所未有的方式融入数字时代的主流呢? 本文将在阐释智慧数据、数字人文及其与图书馆的关系之后,利用一些范例来展示信息服务的新思路,特别针对文本型、非文本型原始数据的结构化和语义化处理的新思路和新方法,证明图档博等机构在语义网和大数据时代不仅是智慧数据的提供者也是直接受益者,智慧数据建设不仅能有效促进数字人文的发展,也将成为图档博机构最重要的新兴工作。

同时,本文特别强调,当把大数据和智慧数据放在数字人文的背景下时,首先要明确“数据”这个术语的含义。在数字时代,人们可能通常认为数据只是数字格式。虽然把数字数据和数据分析联系起来是正确的,但需要充分理解“数据”和“数字数据”的含义不是等价的。数据的类型也不限于定量数据。开放档案信息系统

OAIS 的参考模型 (ISO 14721: 2012 (CCSDS 650.0-P-1.1) Space data and information transfer systems—Open archival information system (OAIS)—Reference model) 将数据定义为“以适合于交流、阐释、处理的形式化方式对信息的可重新解释的表示”,同时提供了数据的示例:比特序列、数字表格、页面上的字符、讲述者的声音记录、亦或月球岩石样本。“数据”这个定义是在“信息”的语境下给出的,信息是“任何可以交换的知识类型。在交换过程中,以数据为表现形式”^[1]。Borgman (2015) 在其《大数据,小数据,无数据:网络世界中的学术研究》一书中曾对“数据”的定义和相关术语进行了全面回顾,进而提出一个总体概括:“数据是对用于学术研究的有关某现象的观察结果、事物对象,或其他作为现象的证实的实体的表现形式”^[2]。基于这样的定义来讨论图档博数据资源的巨大价值,可有效引导我们思考如何运用数字人文手段来挖掘这些无价之宝。

1 从大数据到智慧数据

1.1 智慧数据的定义

在刻画大数据的特征时,往往可以见到多个“V”,而这些“V”还在不断增加。除了数据的规模 (Volume)、数据的流转速度 (Velocity) 和数据的类型多样性 (Variety) 外,还有其他维度,如数据的易变性 (Variability) 和数据的真实性 (Veracity) 等。在合理使用的情况下,大数据可以带来另一个最重要的“V”:巨大的价值 (Value)。通过对带有这些特征的大型数据集的有效处理,可以从中发现隐藏的模式、意外的相关性和令人惊讶的联系^[3]。而“智慧数据”就是实现大数据特征中最后一个“V”——价值 (Value) 的方法,即通过对任何规模的可信的、情境化的、相关切题的、可认知的、可预测的和可消费的数据的使用来获得重大的见解和洞察力,揭示规律,给出结论和对策^[4]。简单来说,智慧数据就是从大数据中得出有意义的信息^[5]。智慧数据的

价值是在大数据的容量、速度、多样性和真实性基础上,通过提供可操作的信息和完善决策来实现的。智慧数据代表着通过多源数据(包括

大数据)的融合、关联和分析等活动实现决策辅助和行动的方法^[6]。图1展示了大数据与智慧数据之间的递进关系^[7]。



注: 图片来源: Zeng, 2017. [7] Compiled based on Kobielus, James. (2016, June). The Evolution of Big Data to Smart Data. Keynote at Smart Data Online 2016

图1 大数据与智慧数据

从情境化、可认知、可预测的特点来看,智慧数据通常带有自描述机制,背后有领域本体作支撑,使得这些数据符合特定的逻辑结构和形式规范,而且支持推理,由此形成智慧的基础,产生可预测和可消费的数据。智慧数据是一种人和机器都能读懂的编码化知识,同时是便于机器理解的数据,而非只有机器可用的、难于表达的概率性隐性知识。智慧数据具有较强的可解释性,支持逻辑推理,这使得它能够用于多种用途和支持多种互操作,并且具有很强的可追溯能力,能够满足人文研究范式的需要。

1.2 智慧数据内涵剖析

21 世纪的数据,如同 18 世纪的石油,对试图提取和利用它的人来说是未开发的资产,拥有不可估量的价值。近些年来,“数据是新的石油”^[8]这一措辞,越来越令人信服。“然而,数据是未开发的原油,它需要经过精炼和加工才能产生真正的价值,需要经过清洗、转换和分析才能释放出其潜能。”^[9]也就是说,数据的原始形

态好比原油,需要进行提炼与处理,才能变成有用的能源。2012 年的一份“数字宇宙”报告显示,有标记的数据仅占数字宇宙的 3%,其中以分析的仅占 0.5%^[10]。由此可见,从带有其他“V”特征的大数据中挖掘价值,面临着巨大挑战和机遇。

数据必须被清洗、转换和分析,以释放其潜力,一旦经过组织和整合,大量的非结构化、半结构化以及结构化数据将变成能反映特定学科或领域研究重点的“智慧数据”。这些智慧数据,可以用作综合分析并产生新的产品与服务^[3,9,11-12]。智慧数据本意并不是说数据具有智慧,智慧是人类的能力。人们从数据中发现新知的应用智慧,是一种能力或者说是生产力。通俗来讲,从大数据中提炼智慧数据,关键不在于你拥有多少数据,而是在于你如何使用这些数据,如何更好地让数据发挥作用。

1.3 智慧数据的现实解读

智慧数据会议(Smart Data Conference)是一

个多学科交叉融合的数据技术盛会。通过对 2015—2017 年的智慧数据会议的主题进行整理,可以识别出与实现智慧数据相关的技术。这些技术主要包括:认知计算、深度学习、机器学习、人工智能、预测分析、图数据库、机器智能、语音处理、语义技术、自主载体、大数据、数据科学、物联网、文本分析、资源描述框架(RDF)、知识图谱、情境计算、关联数据、深层因果推理、本体、JSON-LD(一种轻量级的关联数据格式)、常识(Common Sense)、自然语言处理、语义搜索等^[13]。上述主题是紧密关联且相互重叠

的。比如,深度学习在自然语言处理中显示出巨大潜力;认知计算利用机器学习在复杂的、非结构化的、流式的数据中发现深层模式(包括那些不明显为统计数据的)。一些主题已经跨越其原领域范畴,例如“人工智能”就是一个在 21 世纪发生了巨大变化的领域。同时,2017 年智慧数据会议的主题还反映出 W3C 语义网标准的各种应用,包括(但不限于)RDF、关联数据、本体、图数据库、语义搜索和其他语义技术,如图 2 所示(根据 2017 年“智慧数据”大会日程(包括同场议题)整理)^[13]。

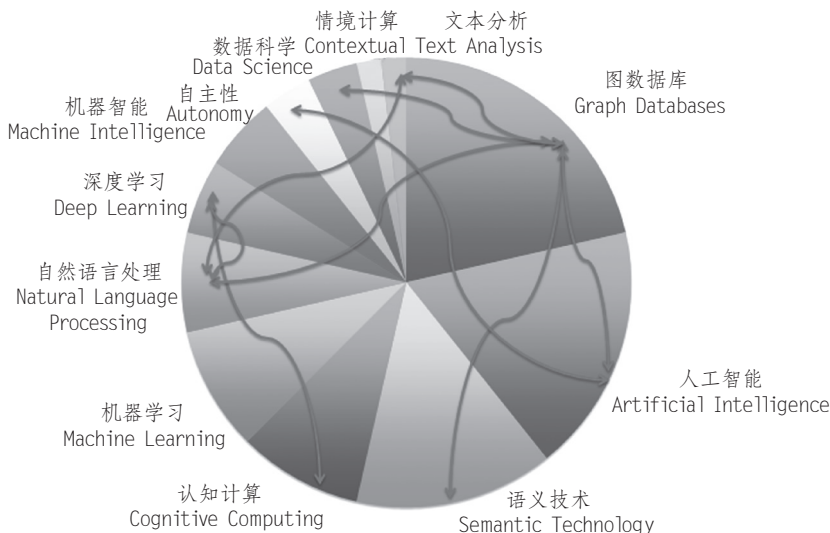


图 2 2017 年“智慧数据”会议议题分析

2 智慧数据应用于数字人文领域以及与图档博数据服务的关系

2.1 从数字人文领域研究项目观察发展趋势

在人文研究领域,尽管“智慧数据”这一词还未被广泛使用,但在过去的六年里,智慧数据的方法却已经被许多研究项目所认可。表 1 整理了自 2009 年以来“数据挖掘挑战”研究计划(<https://diggingintodata.org/awards>)的立项信息,来自十多个国家的研究资助者已资助几十个项目,旨在研究人文和社会科学领域的问题。

美国方面的赞助者主要包括美国人文基金会(NEH)、美国国家科学基金会(NSF)、美国博物馆和图书馆服务协会(IMLS)。通过分析过去四轮计划的摘要可以发现,数据资源既包括以往的非结构化数据资产,也包括数字时代的结构化数据。这些资源在人文社会科学的相关领域里广泛传播。技术上,大规模数据分析已经借助智慧数据方法在相关研究领域得到应用;研究方法上,该计划是跨学科性的,并致力于通过分析大规模和不同格式的数据来挖掘研究重点,但同时也要确保人文和社会科学研究者可以利用新技术工具使用这些数据,如表 1 所示。

表1 2009—2016年“数据挖掘挑战”(Digging into Data)的立项信息汇总
(列表按英文字母排列)

Domains/Areas of Interests 参与的专业领域	Resources 资源	Approaches 方式
activities in humanities & social science (人文与社会科学活动)	audio (music) recordings (音频记录)	annotation (标注)
ancient language (古代语言)	cuneiform tablets (Mesopotamia) (楔形文字泥板)	comparative analysis (对比分析)
archaeology (考古学)	folklore collections (寓言故事集)	computational analysis (计算分析)
biodiversity (生物多样性)	GDP per capita (人均国民生产总值)	computing (计算)
child language development (儿童语言发展)	geographical data (地理信息数据)	corpus building (语料库构建)
colonisation of America (美国殖民史)	GitHub (代码仓库)	cross-datasets analysis (跨数据集分析)
comparative and epidemiological paradigm (比较及认识论范式)	journals (期刊)	cross-datasets searching (跨数据集搜寻)
criminal intent (犯罪意图)	knowledge graphs (知识图谱)	cross-linguistic annotation (跨语种标注)
debating (论辩)	knowledge organization systems (知识组织系统)	data management (数据管理)
early modern common placing (近代摘录备忘)	letters (书信)	data mining (数据挖掘)
economics (经济学)	linguistics databases (语言学数据库)	image processing (图像处理)
English speech (英语演讲)	manuscripts (手稿)	indexing (索引)
epidemiology (流行病学)	manuscripts (pre-modern European) (手抄本(近代欧洲))	linking (关联)
film and media history (电影及媒体历史)	maps (地图)	machine coding (机器编码)
financial system (财政系统)	medical images (医学图像)	machine learning (机器学习)
history (历史)	medieval charters (中世纪宪章)	machine translation (机器翻译)
human migration (人类迁移)	multilingual classic text (多语言古典文本)	metadata aggregation (元数据聚合)
human rights violations (人权侵犯)	music info (音乐信息)	metadata analysis (元数据分析)
information networks (信息网络)	news about terrorism (恐怖主义消息)	metadata auto-generation (元数据自动生成)
information patterns and behaviors (信息模式及信息行为)	newspapers (报纸)	metadata extraction (元数据抽取)

续表

Domains/Areas of Interests 参与的专业领域	Resources 资源	Approaches 方式
journalism (新闻学)	open access publications (开放获取出版物)	natural language processing (NLP) (自然语言处理)
language evolution (语言进化)	papyrus documents (莎草纸文献)	protocols development (协议开发)
legal structures (法律结构)	passages (文章片段)	spatial-temporal correlation (时空关联)
linguistics (语言学)	poetry (陶器)	speech mining (话语挖掘)
literary networks (文学网络)	population databases (人口数据库)	text analysis (文本分析)
manuscripts provenance (手稿由来)	proceedings (会议录)	visualization (可视化)
music (音乐)	quotations (引述)	
musicology (音乐学)	records in indigenous style (固有风格记录)	
parliaments (议会)	records in Spanish (西班牙语记录)	
policy (政策)	signs (符号)	
population (人口)	social media (社交媒体)	
railroad (铁路)	speech datasets (演讲数据集)	
social science (社会科学)	speech recordings (演讲录音)	
sociological theory (社会学理论)	speeches (演讲)	
standards of living (生存标准)	spoken language collections (口语集合)	
storytelling traditions & story repertoires (叙事传统及故事库)	tweets, political (推特,政治因素)	
trading and financial markets (贸易及金融市场)	video data (视频数据)	
vocabularies (词汇)	writing pieces (写作片段)	

最近,美国人文基金会(NEH)举办了一项全国性比赛,鼓励参赛者使用来自《记录美国》(Chronicling America)数字仓储中具有历史意义的美国报纸数字资源和由美国人文基金会资助的新的人文科学研究计划^[14],这意味着人文学科和数字技术的交集不仅走向深入而且更加普及。

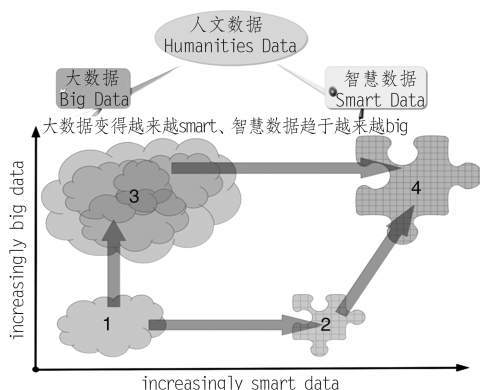
虽然我们尚未充分理解数字人文的多面性,但是我们还是可以从如火如荼的数字人文

国际会议(Digital Humanities 会议,简称 DH 会议)中发现更多线索。通过分析 DH2013 至 DH2016 年会议的主题标签,可以看到数字人文具有多学科属性:文本分析排在首位,其次是历史研究、数据挖掘、文本挖掘、档案库、文学研究和数据可视化。DH2017 年会议将主题和学科进一步分离,跨学科合作和资料库操作十分明显。其中超过 100 篇论文的学科有:计算机科学、文学、图书情报学、文化研究和历史学。一

个值得注意的发现是,电影和传媒领域与其他非文本型专业的研究论文大大增加,而且新进作者数量和作者合作论文数量也呈稳步增加趋势^[15-16]。

2.2 数字人文中体现的智慧数据与大数据方法

根据过去六年里政府资助的研究项目、数字人文会议展示的成果以及世界各地的新举措和学术出版物等信息来观察,人文学科领域内实现“更大的智慧数据”或“更智慧的大数据”的方法已广泛存在^[12],图3所表现的正是这种方法论的阐述,即任何原始数据(图中#1)均可以向智慧数据方向发展,使之结构化(图中#2),同时也可向大数据方向延伸,形成更大的数据(图中#3),结果是:大数据变得越来越智慧,智慧数据趋于越来越大(图中#4以及水平与垂直两个坐标方向)。



注：根据Schöch(2013)文章(CC-by)[12]加工。

图3 “更大的智慧数据”或“更智慧的大数据”

当我们在数据密集型研究项目中考虑数字人文时,人们可能会寻找其与技术相关的独特标志。然而,Schöch(2013)指出大数据在人文领域的独特标志在于方法论的转变,而非技术^[12]。人文研究方法论的转变进一步强调了知识领域的大数据和智慧数据的作用^[17]。大数据转化为智慧数据的视图要追溯到著名的数据(Data)—信息(Information)—知识(Knowledge)—智慧

(Wisdom)的DIKW金字塔^[18-19],它代表了理解一个远远超过我们大脑能力的世界的最基本策略,即过滤、筛选,或者将其精简为更有意义的东西,从数据上升到智慧。然而,智慧数据的实现方法不是简单的复制DIKW路径,因为智慧数据是基于大数据的方法,即为了揭示“未知—未知(the Unknown-Unknown)”而采取的方法^[20],而非为了证明或否定“已知—未知(the Known-Unknown)”,这是智慧数据区别于其他遵循传统蓝图(即假设、建模和测试的方法)的根本所在^[21]。

在《科学》杂志和《自然》杂志的录像“Nature Video”上发表的研究项目“文化历史的网络构架”(A Network Framework of Cultural History)是通过智慧数据揭示“未知—未知”的一个极好的案例^[22-23]。一个多学科研究团队通过处理约15万个著名人物出生和死亡时间、地点这样简单而大型的数据集,展现以往无正式文件证明的人口移动和文化迁徙模式。通过数据分析,跨越三千年的欧美文化模式被可视化重现,揭示知识文化中心的形成及相应的帝国兴衰和其他社会经济技术等因素,为研究欧洲和北美文化历史提供了一个宏观的视角,超越了就事件本身或有限时间范围做研究的方法,其知识的价值是惊人的。而这巨大的价值正是通过对可信的、情境化的、相关且切题的、可认知的、可预测的和可消费的数据的使用而揭示出的深刻见解。该研究所采用的原始资料来自Freebase网站(现为维基数据 Wikidata)、艺术家辞典(Allgemeines Künstlerlexikon, AKL)和艺术家联合人名规范文档(Union List of Artist Names, ULAN)中的结构化数据。

这个案例不仅大体上展示了智慧数据方法在社会学、人类学和历史学中的潜力,还表明了人文研究领域方法论的重要进步,可总结为:大数据的模式和思维方式、智慧数据的实现方式;关注“是什么”,不追究“为什么”;揭示“未知—未知”;将非结构化数据转变成半结构化数据和结构化数据,使大数据变得更智慧。

2.3 数字人文与图档博的角色

数字人文研究和智慧数据的使用都与图档博机构紧密相关。表1中“数据挖掘挑战”过去四轮计划的摘要信息表明,图档博机构所拥有的数据包括很多非结构化数据资源,它们对数字人文研究来说是无价之宝。2016年12月,美国国会图书馆召开了文化遗产与数据国际研讨会(Cultural Heritage & Data)^[24]。该会议旨在跨欧美分享有关文化遗产研究的成果,探讨研究基础设施的作用。会议组织者包括欧洲文化遗产研究基础设施综合平台(IPERION CH)、欧盟国家文化研究所、美国国家美术馆、史密森博物馆保护研究所、乔治华盛顿大学、美国国会图书馆等。来自欧洲和美国的专家探讨了文化遗产领域目前正在实施的研究、教育和培训活动,文化遗产分析和诊断仪器与方法,以及数字遗产相关问题。如果没有图档博的庞大原始材料和数据资源做支撑,文化遗产学科及数字人文研究的深入则不可想象。

2015年底,美国图书馆协会曾就数字人文与图书馆角色进行过一次调查^[25]。其中,图书馆员的调查显示:97%的受访者认为数字人文的资料和项目成果应纳入图书馆馆藏;51%的受访者认为在项目起步时提供咨询帮助是图书馆员帮助数字人文用户的一个重要方式;41%的受访者认为他们的数字学术服务是特别指定(ad hoc)的;21%的受访者表示他们创建了特殊职位(如“数字人文图书馆员”),其他的则将现有员工培训成为项目合作者。大学教师的调查显示:几乎所有受访者都认为对数字人文的支持提高了大学图书馆的重要性;60%的受访者认为数字人文中心属于图书馆数字资源中心;一半以上的教师认为图书馆不仅是机构仓储,还有推动相互配合的数字化服务,将各种服务有机结合,提供学术上使用的元数据等作用;教师们通过其自身的工作,体会到图书馆的任务包括总体支持、服务中介、工具培训、查找资源等。他们将图书馆视为一个数字中心,证实图书馆作为研究基地,显示出图书馆员专业技能的价值。

3 智慧数据方法在图档博数据中的可应用性

本文在前言中已经谈到“数据”这个术语的使用不应只限制在数字数据,数据是“对于学术研究的有关某现象的观察结果、事物对象、或其他作为现象的证实的实体的表现形式”^[2]。

3.1 图档博数据的常见类型

数字人文研究的数据通常来自于图书馆、档案馆、博物馆以及其他信息机构,这些机构为人文科学研究者从结构化数据和半结构化数据中挖掘价值提供了巨大的机会。图4展示的是常见图档博数据资源的实例^[7]。

结构化数据包括书目、索引及文摘数据库、引文索引、各种类型的目录、专藏门户和导航式指南、元数据注册和存储库、归档研究数据集、名称规范档和知识组织系统等。半结构化数据包括根据TEI(Text Encoding Initiative)编码的经典作品、档案找寻辅助工具(Finding Aids)、以各种形式存在的增值的或标记过的资源,以及元数据记录中所包含的非结构化部分等。这些数据集与大数据相比可能体积较小、异质性有限,但它们是干净的、明确的、受信任的和增值的,它们的产生主要与人类决策有关。它们最有可能存在于自由开放的资源和学术资源中(非专属和非商业),这些均为所有人文研究学者所珍视。Richard Wallis是图书馆界关联开放数据运动中的著名先锋,在其2016年智慧数据在线会议的演讲《伴随知识图谱和实体网络的语境运算》中,提到了他对语境运算(Contextual Computing)的愿景,诸如语意、语法、时间、地点、合适领域、规则、用户资料(User Profile)、过程、任务和目标等元素。其中成功的案例有OCLC旗下的图书馆馆藏全球目录WorldCat的关联数据和WorldCat实体(WorldCat Entities)项目。WorldCat实体通过提供成千上万的实体,包括作品、地点、概念、人物、组织、事件和其他类型的处理数

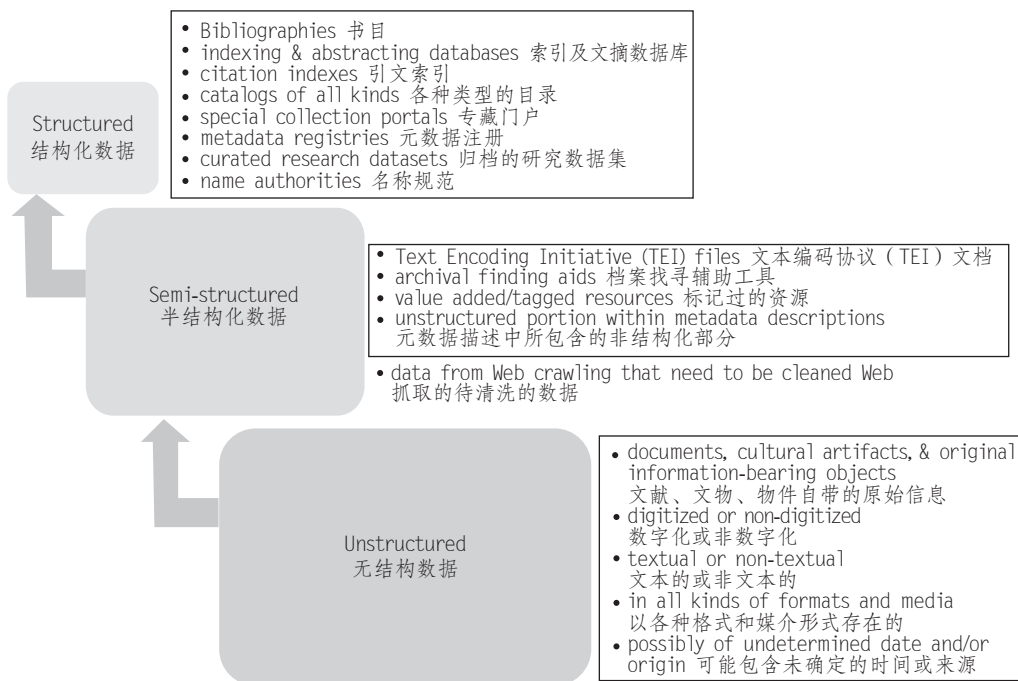


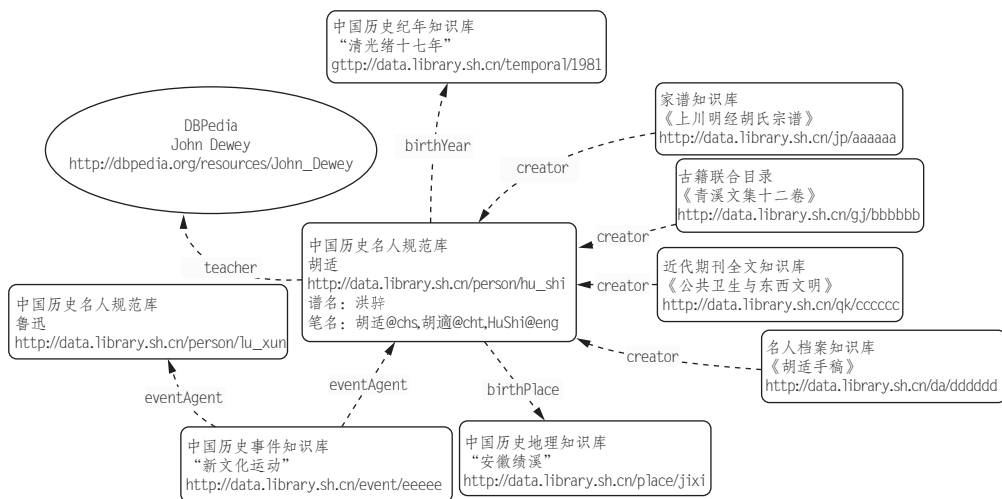
图4 图档博数据示例 (LAM Data Example)

据,展示了由图档博提供的结构化数据如何能无限地丰富知识图谱与关联数据集^[26]。

然而,仅仅依赖于书目数据等结构化数据还不够,因为传统的馆藏资料库往往不是以人文学科研究的方式构建的,这使得数字人文研究者难以直接使用。为此,近两年来上海图书馆利用馆藏开发了一系列文献知识库。几个已经完成的有盛宣怀档案知识库和家谱知识库。几个基础知识库(历史人物规范库、历史地理知识库、历史纪年知识库、历史事件知识库)则分别对应人、地、时、事这四种与文献内容密切相关的维度,具体见图5^[27]。这些知识库都以关联开放数据(LOD)的方式,解决了处理不同结构和格式的文献的有情境的知识关联问题,并向人文学科研究者提供了面向内容的开放资料服务。加上在资料底层实现各类多媒体文献知识库的互联互通,整个面向数字人文研究的资料基础架构应运而生,如图5所示^[27]。这

种方法在《数字人文的技术体系与理论结构探讨》一文中,正是“内容的智慧化”的具体实现方法。该文指出智慧化通常有两方面的含义:内容的智慧化和服务的智慧化^[28]。具体来说,内容的智慧化有赖于对数据进行结构化编码,提供知识的基本结构单元,通过知识本体的构建和关联关系的计算而模拟人脑的认知过程进行前端“机器学习”,而在后端达成一定的“智慧”。

对图档博机构来说,那些量最大、类型最复杂、性质和质量不一的数据,是资源数据化中挑战最大的领域。其中,最难以处理的是文献和其他含有信息的资源中的非结构化数据,不管是文本还是非文本的、数字化的还是非数字的,其形式可说是千奇百怪,具体可参考表1中关于各项目采用的数据类型。这些重要的数据资源通常被保存在特藏、档案、口述历史资料、年度报告、采购与收藏记载索引、库存记录等资料类型里面。很明显,这些数据的性质完全不同



注: 图像来源: 夏翠娟(2017)文章 (CC-by) [27]

图5 上海图书馆文献知识库人、地、时、事与多种类文献关联关系示例

于本文第一章提到的“数字宇宙”中使用的大数据——“从移动电话上传到 YouTube 上的图像和视频、高清数字电影、ATM 的刷卡信息、机场和奥林匹克运动会等重大活动的安全录像、CERN 的大型强子对撞机记录的亚原子碰撞、记录高速公路收费的转发器、通过数字电话线路的语音呼叫以及作为普遍通信手段的短信”等^[10]。显而易见,这样的“数字宇宙”往往不是人文研究的主要数据源,图档博服务在融入大数据潮流的同时应明确认识到这一点。

在数字人文领域应用智慧数据方法的挑战首先在于如何找到历史性数据源,因为这些数据无法利用网络爬虫获得。在商业和工业中测试和实施的智慧数据方法可以应用于数字人文领域,但在学者们分析数据之前,如何将非结构化数据“数据化”?例如,如何将文化遗产材料转化为不仅是机器可读的而且是机器可处理的资源,并通过数字化流程进行重建?这个问题可能解释了为什么在数字人文领域,智慧数据方法强调通过组织和整合过程来将非结构化数据转换为结构化和半结构化数据^[12,29-30]。在将非结构化数据转化为结构化和半结构化数

据过程中,智慧数据提供者的服务旨在获得机器可读的、机器可处理的和机器可操作的(而不仅仅是机器可读的)数据,以提供用于链接、引用、转移、权利许可管理、使用和重复使用的精确数据,同时实现数字人文领域中一对多功能和高效数据处理^[26]。这些正是图档博,特别是大学图书馆和研究型图书馆近年来的主要努力方向之一。

3.2 非文本数据中的图像数据

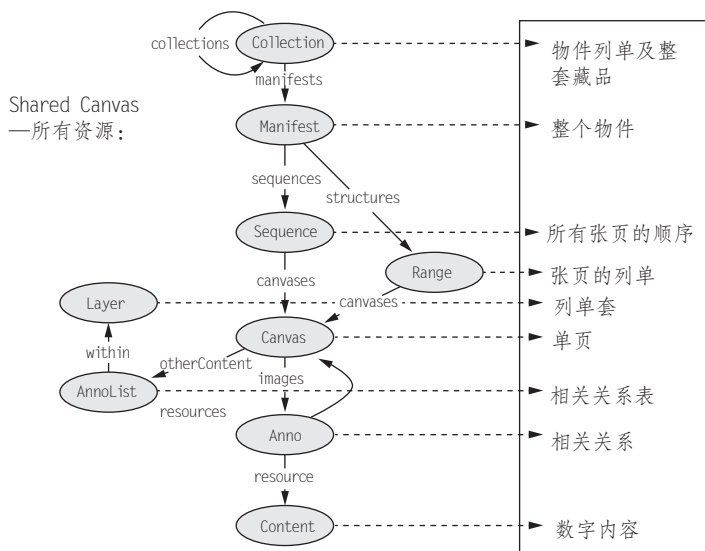
文本型非结构数据的处理技术近年来日趋成熟,在许多项目中都有报道,也是数字人文大会中发言者提到次数最多的领域之一。相比之下,对于非文本型的数据资源来说,智慧数据的应用和前途在何处?图像,作为一种传递信息、知识和思想的视觉媒介,能够表示文本难以表达的复杂信息。图像的具体表现形式十分多样,包括绘画、照片、草图、手稿等。在文化遗产领域,图像常常以壁画、油画、织锦画等形式出现,包含了深刻的文化内涵,复杂的时空场景和较为抽象的思想寓意。对一件实物作品而言,数字化之后往往会产生多个图像。在此我们特意针对图像的数据化操作和深度语义化处理进行讨论。

3.2.1 数字图像互操作框架

美国和欧洲顶尖大学的图书馆、国家图书馆、博物馆联合努力四年,制定了《国际图像互操作框架》(International Image Interoperability Framework, IIIF)的四个应用程序接口 API 协议^[31]。该框架旨在解决文化资源数字化后文件难以被发现、再利用、引用、交换、比较分析等难题。多年来,大家都在采用不同的工具和系统开展数字化工作,形成了大大小小的独立数字王国,这些系统不仅昂贵、功能简单,其内容范围也受局限,分散在不同地方的相关物件的图像难以放到一起比较或分析。如果采用标准化的应用程序接口,图像和元数据就可以无障碍交互,图档博也可以在使用层中对不同技术进行选择。

在 IIIF 制定的标准应用程序接口 API 协定中,首先是“图像 API”,具体到对图像的特定区域、位置、大小、角度、材质、格式的统一的句法表现:在图像的获取和交换中 URL 的语义内涵和格式为“{根基}/{区域}/{大小}/{角度}/{质量}.{载体格式}”。其“呈现 API”首先用于

合作性的加注、修正、语义抽取,允许对图像中任何部分进行标注,标注工具、描述工具、地理信息工具均可配合使用。该协议的目的是将来自不同图档博机构的相同图像的多个不同版本进行比较,或将分散在不同馆藏的相关物件放到一起,互相补充,整合为一体。这些应用程序接口 API 协议通过机器可理解的结构化数据来解决问题,其结构化数据的层次在图 6 中有所体现,约定文献对每种资源的元数据要求明确指定必须具备、建议具备、可有可无、不允许等类型,采用尽量简单的格式,避免不必要的操作。IIIF 还进一步开发了“验证 API”和“查寻 API”。许多世界顶尖图档博的图像管理系统已经采用了这些协议,服务功效十分显著,如:促进丰富多彩的图像递送;可以用在原有系统上,不管是服务器还是客户端、插件等;出版一次,复用多次;将众人提供的资源重新组合;可由使用者直接进行标注;全球性的支持和内容网络;支持获取途径控制、归因;促使有出处来源和受控的分享等^[32-33]。



注: 图片根据IIIF呈现API [http://iiif.io/api/presentation/2.0/\(CC-by\)](http://iiif.io/api/presentation/2.0/(CC-by))编译

图6 《国际图像互操作框架(IIIF)》结构化数据的层次

3.2.2 图像的深度标引

近年来,随着文化遗产数字化加速,文化遗产领域的数字图像资源数量激增,由此激发的图像相关的数字人文研究反过来又对图像数据资源建设提出了更高的要求。传统的图像资源组织主要借助元数据和大众标注(Folksonomy)实现其语义化描述。元数据和大众标注主要关注于图像的外部特征,如拍摄人、拍摄时间、拍摄地点等。关于图像主题(subject)的描述和揭示也主要关注图像的关涉物(aboutness),而图像内部对象(ofness)的描述和揭示往往被忽略,由此导致图像资源的语义化加工深度不够,难以实现图像细粒度内容的深度集成和认知计算。

要实现原始图像数据向智慧数据的转变,需要对数字图像进行深度语义组织,即深度标引(Deep Indexing)和形式化语义表示^[34-35]。图像深度标引的目标是要深入到图像内部,关注图像内含的各种片段语义信息和整体语义信息,即对象语义信息和状态语义信息(ofness和isness)。以敦煌壁画为例,深度标引不仅要详细标引画面中存在的佛像、动物、交通工具等实体对象类型,还要指明它们具体的名称,如释迦摩尼、九色鹿、步辇等。对于复杂的叙事型图像,深度标引不仅要把图像内的具体人物和器物揭

示和标示出来,还要描述和表示事件的时空与情节^[36]。如果图像中包含多个情节,每个情节的参与人物不同,那么深度标引还要明确标示情节与各个实体对象以及行为动作的关系,以及行为动作之间的关系^[37]。而已往的图像元数据模型难以满足这样的标引要求。

图像深度标引过程既可以全部由专家实现,也可以由大众协同完成,还可以利用计算机辅助实施。在自动图像深度标引中,图像分割(Image Segmentation)技术可以确定待标引对象,并将其从整副图像中识别和分割出来。然后,图像语义标注(Image Semantic Annotation)技术和图像自动说明(Image Caption)技术可以实现分割后对象的分类和精准识别,以及自然语言描述。图像深度标引信息需要与图像内的特定区域,也就是分割好的画面片段相对应,如图7中右侧部分的方框部分、动物形状部分和人形部分,分别对应着左侧的“溺人拜恩”“九色鹿”“溺人”三个实体,分别是情节实体、动物实体和人物实体。这些实体之间带有特定的逻辑关系,如图7中左侧部分的树状结构所示^[35]。图像深度标引可以借助开放协同标注(Open Annotation Collaboration, OAC)和国际图像交互框架IIIF实现标准化表示,同时也可以表示成RDF三元组的形式^[37]。

Image pols Tours Concepts Search

Points of Interest within this image
List Hierarchy Expand All Collapse All

- 01. 鹿在河边散步
- 02. 鹿救溺人
- ▼ 03. 溺人拜恩
 - 03.01 九色鹿
 - ▼ 03.02 溺人调达
 - 03.02.01 披巾
- 04. 王后说梦
- 05. 溺人告密
- 06. 众人捕鹿
- 07. 乌鸦叫醒九色鹿
- 08. 九色鹿怒斥溺人



图7 图像深度标引示意图

图像深度标引不仅揭示了图像的主题 (Subject), 描述了图像内容, 还在一定程度上实现了图像从非结构化的、难于理解的数据形态向结构化的、便于认知计算的数据形态的转变。这些语义化表示的标注信息与图像资源本身一同构成了图像智慧数据资源, 对语义检索、细粒度内容资源集成、知识发现、跨模态认知计算、深度学习等计算机处理形成了资源支撑, 有利于构建面向数字人文研究的人文智慧数据基础设施。

3.3 非物质文化遗产数据

非物质文化遗产 (简称“非遗”) 是有温度的活态文化, 既蕴涵了历史传承的文化基因, 又体现了与时俱进的时代气息。非遗的保护是抢救式与开发式并举的, 数字化是手段, 文化传承与创新是核心精神所在。数字人文为非遗数字化保护提供了文化创新的新理念、方法与技术, 也是开启非遗智慧数据大门的一把钥匙。

中国自 2004 年加入联合国《保护非物质文化遗产公约》以来, 非物质文化遗产保护工作取得了巨大成就。截至 2017 年, 中国在国际上申报成功的联合国非遗项目共有 39 项, 其中亟需保护的非物质文化遗产名录项目 7 项, 人类非物质文化遗产代表作名录项目 31 项以及非物质文化遗产优秀实践名册项目 1 项^[38]。在国家层面, 目前已自上而下建立了“国家—省—市—县”四级保护体系, 先后公布了四批国家级非物质文化遗产名录, 共计 1 327 项^[39]。在省市县各级的非遗保护基层工作中, 非遗项目总体数量庞大、文化内容丰富、形态多元。

在非遗普查与保护进程中, 非遗保护工作者通过田野调查、实地走访等搜集、整理、归档了大量资料, 文本、图像、音视频等各种类型的数字资源数量已经达到了大数据级别。非遗项目的数字化基础工作卓有成效。接下来, 在资源“量”的基础上, 如何深入挖掘资源的“质”, 真正发挥非遗的人文魅力, 创造文化价值, 则需要专业的信息资源管理作保障, 通过数字人文方法的探索应用来实现。

以智慧数据要求与数字人文的结合应用为出发点, 非遗数字资源管理与开发目前存在三个主要问题: 数据揭示问题、数据分类问题与数据开放问题。

非遗的数据揭示问题主要指元数据描述的标准化。非遗项目种类多样, 涉及的主题与表现形式多样, 这使得统一元数据标准的制定难度大, 行业实际工作中存在接受障碍。数字人文提倡数据的结构化与语义化基础, 一种可能的解决思路是: 以非遗项目申报管理与普查工作为主体, 设计非遗顶层元数据方案, 奠定基础数据属性的揭示; 在此基础上, 分类别设计非遗项目的元数据应用纲要, 扩展各类别文化特征维度。另外, 考虑到数据共享与互操作需求, 可以考虑复用 DC 元数据及限定词, 参考国际文物和博物馆界联盟组织发布的 CIDOC Conceptual Reference Model (CRM)、Cataloging Cultural Objects: A Guide to Describing Cultural Works and Their Images (CCO)、Visual Resources Association Core Categories (VRA Core) 等相关概念模型与元数据标准。

非遗的数据分类问题一直以来讨论较多, 科学合理的分类体系是构建非遗知识图谱的基础要件。但目前尚未有一套详尽的非遗分类体系标准。联合国教科文组织《保护非物质文化遗产公约》中的非遗内容包括五个方面, 将文化场所也作为代表作名录的一个类别, 合计 6 类^[40]。中国的非遗名录体系由 10 个类别组成。非遗保护实际工作中曾制定《非物质文化遗产分类代码表》, 采用二级分类, 16 个一级类目^[41]。非遗文化的多元特征使得单一的线性分类无法满足多维度揭示与知识检索需要, 另外一些非遗类别的复合主题也需进一步细分优化, 例如, 传统手工艺类和民俗类就有细分的需要。在非遗保护工作的实践基础上, 一种可能的解决思路是: 结合非遗名录建设基础, 编制非遗分类主题词表, 采用分面分类与叙词表结合的方式, 重视多维揭示与概念组配, 既能满足业务工作的数据资料管理需要, 又能满足非遗知

识获取、关联与发现的服务要求。

非遗的数据开放问题是资源长期累积之后所产生的现实问题。非遗数据目前停留在基本网页发布(半结构化数据)与基础数据库(结构化数据)的初步检索阶段,轻量级碎片化数据多,高质量的非遗开放数据集缺乏。可通过语义建模、关联数据发布与知识可视化等技术,采用开放数据格式标准,如 RDF、JSON-LD 等,提升非遗项目的数据开放质量。在数据开放的同时,加强非遗之间的关联性,探索各种显式与隐式关系,例如,非遗项目之间的演化和传承关系,非遗项目与传承人的关系等,这些对了解非遗文化的活态流变性非常重要。

除了以上问题,非遗的数据管理与保障体制也是非遗智慧数据的必备条件。文化部主管的博物馆、非物质文化遗产中心、非遗保护单位(地区)、地方文化馆应加强与图书馆、档案馆的业务交流合作,巩固“技术观”、加强“资源观”、创新“文化观”。

总的来说,现阶段非遗的数字化保护强调归档、公告、报道等基础事实类数字资源类型,非遗资源数据已经具备了“量”的基础,接下来要提升“质”的要求,即从非遗大数据到非遗智慧数据的上升。未来通过数字人文的方法、技

术与手段,推进非遗数字资源的立体化、多维化、动态化建设,达到智慧数据的要求,满足智慧博物馆、“互联网+文化场景”的服务需要。

4 结语

如今,借助大数据和智慧数据以及前沿技术,人文领域的研究人员正以前所未有的技能融入数字时代的主流:访问和再利用海量的多样化数据,发掘以前隐藏的模式和关系,重现过去,在现实和虚拟环境中定性与定量分析影响和价值等。图档博和各种文物机构所拥有的数据是无价之宝,如果采纳大数据的模式和思维方式、智慧数据的实现方式,以非结构化数据到结构化数据的组织和整合过程为主要手段,产生机器可理解的、一源多用、高效率运作的数字,将促使图档博以及相关行业携带这些丰富的资源进入数字时代的主流。虽然挑战与机遇并存,但可以肯定的是,智慧数据,即通过对任何规模的可信的、情境化的、相关切题的、可认知的、可预测的和可消费的数据的使用来获得重大的见解和洞察力,能够产生无法比拟的价值,并会促进数字人文领域的前进与变革。

参考文献

- [1] Consultative Committee for Space Data Systems. Reference model for an open archival information system (OAIS): recommended practice (CCSDS 650.0-M-2; Magenta Book Issue 2. June 2012) [M/OL]. Washington, DC: Management Council of the Consultative Committee for Space Data Systems (CCSDS), 2012 [2017-10-08]. <http://public.ccsds.org/publications/archive/650x0m2.pdf>.
- [2] Borgman C. Big data, little data, no data: scholarship in the networked world [M]. Cambridge, MA: MIT Press, 2015:28.
- [3] Gardner D. An ocean of data [M]//Smolan R., Erwit J. The human face of big data. Sausalito, CA: Against All Odds Productions, 2012:14-17.
- [4] Kobielus J. The evolution of big data to smart data [EB/OL]. Keynote, Smart Data Online 2016. (2016-07-13) [2017-10-08]. Video available at: <http://www.dataversity.net/big-data-smart-data-big-drivers-smart-decision-making/>.

- [5] Sheth A. Transforming big data into smart data; deriving value via harnessing volume, variety and velocity using semantic techniques and technologies [C]. 30th IEEE International Conference on Data Engineering (ICDE) 2014, DOI: 10.1109/ICDE.2014.6816634.
- [6] Iafate F. From big data to smart data [M]. London: ISTE Ltd., and Hoboken, NJ: John Wiley & Sons, Inc. 2015.
- [7] Zeng M L. Smart data for digital humanities [J/OL]. Journal of Data and Information Science, 2017, 2 (1) : 1-12 [2017-10-08]. <https://doi.org/10.1515/jdis-2017-0001>.
- [8] Palmer M. Data is the new oil [EB/OL]. (2006-11-03) [2017-10-08]. http://ana.blogs.com/maestros/2006/11/data_is_the_new.html.
- [9] TiECON East. Data is the new oil [EB/OL]. (2014-05-30) [2017-10-08]. <http://www.tieconeast.org/2014/big-data-analytics>.
- [10] Gantz J, Reinsel D. The Digital Universe in 2020; big data, bigger digital shadows, and biggest growth in the Far East [J/OL]. IDC iView, 2012, 12: 1-16 [2017-10-08]. <http://www.dedupecentral.co.uk/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf>.
- [11] Mukerjee P. Introduction to data science [EB/OL]. (2014-01-12) [2017-10-08]. <http://www.slideshare.net/prithwis/01-intro2-datascienceyantrajalblog>.
- [12] Schöch C. Big? Smart? Clean? Messy? Data in the humanities [J/OL]. Journal of Digital Humanities, 2013, 2 (3) : 2-13 [2017-10-08]. <http://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities/>.
- [13] DATAVERSITY Education, LLC. Smart Data conference (website) [EB/OL]. (2017-01-30 / 02-01) [2017-10-08]. <http://smartdata2017.dataversity.net>.
- [14] National Endowment for the Humanities (NEH). NEH announces the winners of the chronicling America data challenge [EB/OL]. (2016-07-25) [2017-10-08]. <https://www.neh.gov/news/press-release/2016-07-25>.
- [15] Weingart S. Submissions to DH2016 (pt.1) [EB/OL]. (2015-12-07) [2017-10-08]. <http://scottbot.net/HI-AL/index.html?p=41533.html>.
- [16] Weingart S. Submissions to DH2017 (pt.1) [EB/OL]. (2016-11-10) [2017-10-08]. <http://scottbot.net/submissions-to-dh2017-pt-1/>.
- [17] Svensson P. Big digital humanities; imagining a meeting place for the humanities and the digital [M]. Ann Arbor: University of Michigan Press, 2016.
- [18] Zeleny M. Management support systems; towards integrated knowledge management [J]. Human Systems Management, 1987, 7 (1) : 59-70.
- [19] Ackoff R L. From data to wisdom [J]. Journal of Applied Systems Analysis, 1989, 16 (1) : 3-9.
- [20] Borne K. Big data, small world; kirk borne at TEDxGeorgeMasonU [EB/OL]. (2013-06-10) [2017-10-08]. <https://www.youtube.com/watch?v=Zr02fMBfuRA>.
- [21] Anderson C. The end of theory: The data deluge makes the scientific method obsolete [EB/OL]. (2008-06-23) [2017-10-08]. <https://www.wired.com/2008/06/pb-theory/>.
- [22] Schich M, Song C, Ahn Y Y, et al. A network framework of cultural history [J]. Science, 2014, 345 (6196) :

558-562.

- [23] Schich M, Song C, Ahn Y Y, et al. Charting culture. Nature video[EB/OL].(2014-07-31)[2017-10-08].
<https://www.youtube.com/watch?v=4gIhRkCcD4U>.
- [24] Library of Congress. Cultural heritage and data: the role of research infrastructures[EB/OL].(2016-12-05/06)
[2017-10-08].<https://www.loc.gov/preservation/outreach/symposia/heritagedata.html>.
- [25] Varner S, Hswe P. Special report: Digital humanities in libraries[EB/OL].(2016-01-04)[2017-10-08].<https://americanlibrariesmagazine.org/2016/01/04/special-report-digital-humanities-libraries/>.
- [26] Wallis R. Contextual computing-Knowledge graphs & web of entities[EB/OL].(2017-01-31)[2017-10-14].
<https://www.slideshare.net/rjw/contextual-computing-knowledge-graphs-web-of-entities>.
- [27] 夏翠娟. 以关联开放数据服务为基础的数字人文平台建设方案研究[J/OL].图书馆学与资讯科学,2017,
43(1):47-70 [2017-10-08]. <http://jlis.glis.ntnu.edu.tw/ojs/index.php/jlis/article/view/717>. (Xia Cuijuan.
Building a digital humanities platform by using Linked Open Data services[J/OL].Journal of Library and Infor-
mation Science, 2017,43(1):47-70 [2017-10-08]. [http://jlis.glis.ntnu.edu.tw/ojs/index.php/jlis/article/](http://jlis.glis.ntnu.edu.tw/ojs/index.php/jlis/article/view/717)
[view/717](http://jlis.glis.ntnu.edu.tw/ojs/index.php/jlis/article/view/717).)
- [28] 刘炜, 叶鹰. 数字人文的技术体系与理论结构探讨[J].中国图书馆学报, 2017,43(5):32-41.(Liu Keven
W, Ye Fred Y. Exploring technical system and theoretical structure of digital humanities[J]. Journal of Library
Science in China. 2017,43(5): 32-41.)
- [29] Mayer-Schönberger V, Cukier K. Big data: a revolution that will transform how we live, work, and think[M].
New York, NY: Eamon Dolan/Houghton Mifflin Harcourt,2013.
- [30] KaplanF. A map for big data research in digital humanities[J]. Frontiers in Digital Humanities, 2015,2(1):1-7.
DOI 10.3389/fdigh.2015.00001.
- [31] IIIF. International Image Interoperability Framework[EB/OL].[2017-10-08]. <http://iiif.io/>.
- [32] Sanderson R. Introduction to IIIF[EB/OL].(2015-03-16)[2017-10-08].<http://www.slideshare.net/azaro42/introduction-to-iiif>.
- [33] Warner S. IIIF introduction and opportunities at Cornell[EB/OL].(2015-02-05)[2017-10-08].<http://www.slideshare.net/simeonwarner/2015-01-cornellvrvgiiif>.
- [34] Clarke D. Deep image annotation: making a difference in knowledge organization [C/OL]. International Society
for Knowledge Organization (ISKO) - UK. Biennial Conference, London, 2015-07/13/14. (2015-07-07)
[2017-10-08].<http://www.iskouk.org/sites/default/files/ClarkePaperISKO-UK2015.pdf>.
- [35] 王晓光,刘雪梅,夏生平.文化遗产图像深度语义标引方法设计与实现[J/OL].图书馆学与资讯科学,2017,
43(1):98-121 [2017-10-08]. <http://jlis.glis.ntnu.edu.tw/ojs/index.php/jlis/article/download/716/pdf>.
(Wang Xiaoguang, Liu Xuemei, Xia Shengping. Design and implementation of deep semantic indexing on digital
cultural heritage images [J/OL].Journal of Library and Information Science,2017, 43(1): 98-121 [2017-10-
08].<http://jlis.glis.ntnu.edu.tw/ojs/index.php/jlis/article/download/716/pdf>.)
- [36] 李旭晖,吴燕秋,王晓光.基于角色关联方式的叙事型文化遗产知识表示方法[J].图书情报工作,2017,
61(9):116-122.(Li Xuhui, Wu Yanqiu, Wang Xiaoguang. Research on knowledge representation of narrative
cultural heritage based on role-links[J]. Documentation, Information & Knowledge, 2017,61(9):116-122.)

-
- [37] 徐雷,王晓光.叙事型图像语义标注模型研究[J].中国图书馆学报,2017,43(5):70-83.(Xu Lei, Wang Xiaoguang. Narrative image annotation ontology for semantic web[J]. Journal of Library Science in China. 2017,43(5):70-83.)
- [38] UNESCO. The lists of intangible cultural heritage and the register of good safeguarding practices [EB/OL]. [2017-10-08]. <https://ich.unesco.org/en/lists>.
- [39] 中国艺术研究院. 中国非物质文化遗产名录 [EB/OL]. [2017-10-08]. <http://www.ihchina.cn>. (Chinese National Academy of Arts. List of intangible cultural heritage in China [EB/OL]. [2017-10-08]. <http://www.ihchina.cn>.)
- [40] 联合国教科文组织. 保护非物质文化遗产公约 [EB/OL]. (2003-10-17) [2017-10-08]. http://www.crihap.cn/2014-07/02/content_17638153.htm. (United Nations Educational, Scientific and Cultural Organization. Convention for the safeguarding of the intangible cultural heritage [EB/OL]. (2003-10-17) [2017-10-08]. http://www.crihap.cn/2014-07/02/content_17638153.htm.)
- [41] 王文章.非物质文化遗产概论[M].北京:文化艺术出版社,2013:305-330.(Wang Wenzhang. An introduction to intangible cultural heritage[M]. Beijing: Culture and Art Publishing House, 2013:305-330.)
-

曾 蕾 美国肯特州立大学信息学院教授。美国 俄亥俄州。

王晓光 武汉大学信息资源研究中心教授。湖北 武汉 430072。

范 炜 四川大学公共管理学院信息管理技术系副教授。四川 成都 610065。

(收稿日期:2017-10-09)