

数据驱动背景下美国国立医学图书馆研究布局分析^{*}

王 茜

谢秀芳

(中国医学科学院/北京协和医学院
医学信息研究所 北京 100020)

(1 首都医科大学医学人文学院医学信息学学系 北京 100069
2 首都医科大学图书馆 北京 100069)

王 蕾 杨雨生 方 安

(中国医学科学院/北京协和医学院医学信息研究所 北京 100020)

〔摘要〕 基于美国国立医学图书馆近 5 年发布科研战略规划、研究专项及科研资助项目等信息，综合运用统计分析、聚类分析等文献计量学方法技术，从项目实施机制、资源投入、资助类型、资助机构、资助重点等维度研判数据驱动背景下美国国立医学图书馆研究布局动向、特征及经验，为我国医学图书馆和医学信息学科领域科研规划决策提供参考。

〔关键词〕 医疗大数据；医学图书馆；医学信息学；数据科学；科研布局

〔中图分类号〕 R-058 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2023.01.015

A Research Layout Analysis of the United States National Library of Medicine under the Background of Data-driven WANG Qian, Institute of Medical Information, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100020, China; XIE Xiufang, 1Department of Medical Informatics, School of Medical Humanities, Capital Medical University, Beijing 100069, 2Library of Capital Medical University, Beijing 100069, China; WANG Lei, YANG Yusheng, FANG An, Institute of Medical Information, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100020, China

〔Abstract〕 Based on the information of scientific research strategic planning, research projects and funded scientific research projects published by the United States National Library of Medicine (NLM) in the past five years, the research layout trend, characteristics and experience of NLM under the background of data-driven are studied and evaluated from the perspectives of research implementation mechanism, investment, funding type, funding institution and funding focus by comprehensively using bibliometric analysis tools such as statistical analysis and cluster analysis, so as

〔修回日期〕 2022-08-25

〔作者简介〕 王茜，博士，馆员，发表论文 20 余篇；通信作者：方安，博士，研究馆员。

〔基金项目〕 中国医学科学院医学与健康科技创新工程“人口健康科学大数据智能管理与高效利用技术体系建设”（项目编号：2021-I2M-1-057）；中国医学科学院医学与健康科技创新工程“医学知识管理与智能化知识服务关键技术研究”（项目编号：2021-I2M-1-056）；北京市社会科学基金项目“基于科技路线图的脑科学长期发展趋势研究”（项目编号：19GLC055）；国家科技图书文献中心子任务“医学领域规范数据加工和丰富化”（项目编号：2022XM7101）。

to provide references for scientific research planning and decision-making of medical library and medical information science in China.

[**Keywords**] medical big data; medical library; medical informatics; data science; research layout

1 引言

面对数据在加速生物学发现、改善健康及医疗服务进程中日益重要的发展趋势,美国国立医学图书馆(national library of medicine, NLM)积极进行战略转换,立足可信赖生物学资源和健康信息提供者、开放科学倡导者、下一代数据科学家推动者、文化和学术遗产保护者、信息技术创造者及知识发现重要平台的发展定位^[1],将重构与加强医学信息学、数据科学的研究布局列为赋能 NLM 实现生物学发现和驱动健康愿景的优先事项。

自 2017 年以来, NLM 围绕在全球范围内获取、传播生物学信息资源的核心职能^[2],依托馆内和馆外研究计划,先后推出了促进健康数据发现的计算和统计方法研究、大规模生物学研究资产治理方法研究、数据驱动的新型冠状病毒肺炎研究等研究专项,支持健康医学数据被发现、获取、互操作和重用的技术研发,推进大规模医学研究资产自动化和自主化治理、数据智能分析、可视化之外的数据呈现方法、异构数据的语义挖掘、自然语言处理和深度学习、可计算的生物学知识、支持交互式图书馆的“可操作文献”架构等方面的研究^[3],加大医学信息学和数据科学领域研究人才培养力度,致力于成为生物学信息学和健康数据科学领域的领导者和发展典范。

当前,我国医学图书情报机构正处于“十四五”规划的关键布局和实施阶段,面对数智化重塑下的医疗数据治理环境、健康中国战略实施进程中的高质量医学信息供给需求、多源异构的医疗大数据资源、数据密集型科研范式下医学知识供给态势、支撑医学科技自立自强的医疗健康科技信息安全等方面的挑战和机遇^[4],加强医学信息学和数据科学研究布局和顶层设计、打造创新驱动的高质量自主发展的医学信息服务生态体系成为当务之急。鉴于此,笔者基于 NLM 近 5 年(2017—2021 年)发布的科研战略规划、实施的研究专项、资助的科

研项目等信息,综合运用统计分析、聚类分析等文献计量学方法技术,从项目实施机制、资源投入、资助类型、资助机构、资助重点等维度研判 NLM 研究布局动向和特征,总结 NLM 数据驱动背景下的研究布局经验,以期为我国医学图书情报机构和医学信息学科领域科研规划决策提供参考。

2 数据源与分析方法

2.1 数据源

2.1.1 NLM 资助项目 NLM 项目资助数据主要来源于美国国立卫生研究院(national institutes of health, NIH)项目资助开放获取数据库(research portfolio online reporting tools, RePORTER)^[5],该数据库提供 NIH 自 1985 年以来的项目资助信息。论文主要选取 NLM 在 2017—2021 年资助的项目数据,检索策略为:“Funding: Yes” AND “Agency/Institute/Center: NLM” AND “Fiscal Year: 2017, 2018, 2019, 2020, 2021”,检索日期为 2022 年 2 月 20 日,获得 1 286 条项目数据,包括项目名称、申请时间、项目类型、承担机构、项目经费、项目标题、项目摘要、项目关键词等属性信息。由于论文主要聚焦于 NLM 研究布局分析,结合 NLM 资助项目所属类型(activity codes)^[6],主要选取 NLM 在 2017—2021 年间资助的 945 条研究类项目(DP 系列、R 系列、Z 系列、T 系列、F 系列、K 系列)作为分析数据源。

2.1.2 NLM 规划文本 本研究分析依托的 NLM 规划文本包括 2017—2023 年期间的 NLM 国会陈述报告^[7]、2017—2027 年期间的 NLM 战略规划^[4]及项目资助规划文本^[8-9]等。

2.2 分析方法

采用内容分析方法对规划类文本数据进行关键信息提取;采用统计学、文献计量学等方法技术对资助项目数据进行多维度分析。其中,利用 Excel 实现资助金额、项目数量、项目类型、资助机构等方面的统计分析,利用 Python、Gephi 对 NLM 资助

项目涉及的研究、症状和疾病类别（research, condition, and disease categories, RCDC）进行共现网络聚类分析及可视化^[10]，以识别 NLM 研究资助主题分布和重点主题布局。

3 研究结果

3.1 研究推进机制布局（图 1）

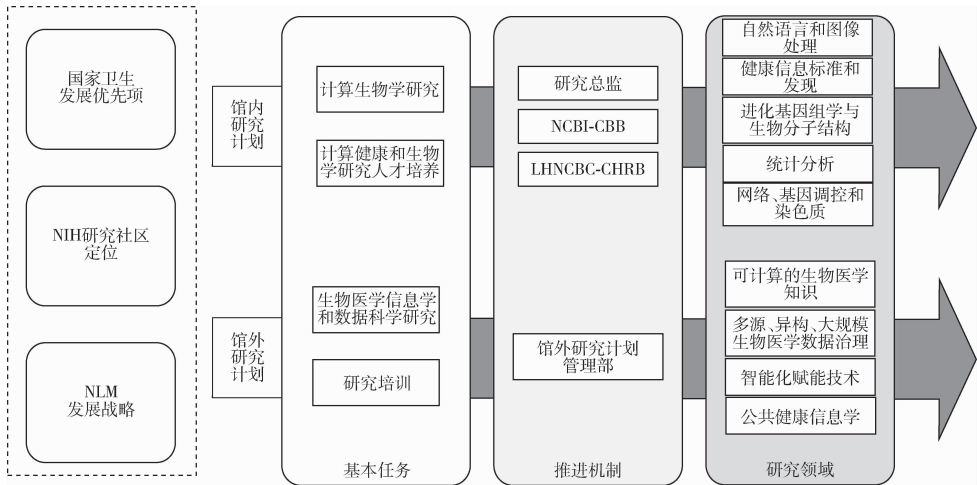


图 1 NLM 研究实施机制布局

NLM 采取馆内研究项目和馆外研究项目双路径驱动的研究推进机制，面向 NLM 下属研究机构及国内外高等院校、大学、医院、小型企业等科技创新主体提供研究项目支持。

3.1.1 NLM 馆内研究计划 NLM 馆内研究计划的具体实施机制为^[8]：聚焦生物医学、分子生物学、健康领域计算方法的研究，面向未来的计算健康和生物学研究人才培养，采取非竞争性资助机制，在科学总监领导下由 NLM 下属的计算生物医学机构和计算健康研究机构负责实施。其中，NLM 研究总监负责馆内研究计划总体方向和质量的监督与管理，美国国立生物技术信息中心计算生物学分部（computational biology at the national center for biotechnology information, NCBI – CBB）负责 NLM 馆内研究计划的计算生物医学研究分支^[11]，Lister Hill 国立生物医学通信中心的计算健康研究分部（Lister Hill national center for biomedical communications, LHNCBC – CHRB）承担 NLM 馆内研究计划的计算健康研究分支^[12]，涉及的研究主题包括网络、基因调控和染色质，统计方法，进化基因组学与生物分子结构，健康信息标准和发现，自然语言和图像处理。

3.1.2 NLM 馆外研究计划 NLM 馆外研究计划的具体实施机制为^[9]：聚焦生物医学信息学和数据科学领域，致力于应用计算机和信息科学来改善生物医学信息的获取、存储、检索、管理、使用等关键问题，在 NLM 馆外研究计划管理部的统筹管理下，面向国内外医学院校、科研机构、企业采取竞争性同行评议的资助方式，提供研究项目和研究培训资助。NLM 馆外研究计划优先考虑创新性高、重要且影响力大的研究议题，如可计算的生物医学知识，大规模生物医学数据库、异构医学知识和多源数据的集成、组织和检索，通过虚拟现实、人工智能和机器学习提高人类智力，医疗决策支持，公共健康和灾害管理相关的信息学研究主题。

3.2 研究投入布局

3.2.1 资助金额和项目数量 2017—2021 年，NLM 资助的研究项目数量和金额整体呈递增趋势。就资助项目数量而言，2017—2021 年间 NLM 共资助各类研究项目 945 项，年均资助项目为 189 项，其中 2020 年项目最多，达到 199 项。在资助金额方面，2017—2021 年间共资助金额 38 775.22 万美元，年均资助 7 755.04 万美元，其中，2021 年资助金额

最多，达到 8 179.22 万美元，见图 2。

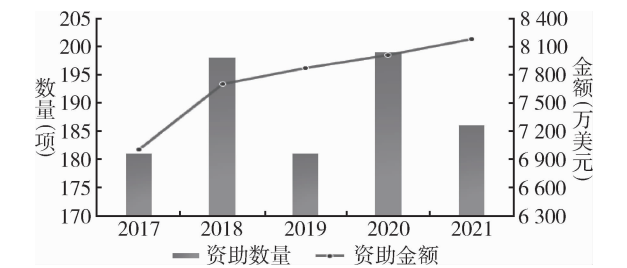


图 2 2017—2021 年 NLM 研究资助金额和项目数量投入概况

3.2.2 资助项目类型 2017—2021 年间 NLM 资助的研究项目类型涵盖研究支持（DP 系列、R 系列和 Z 系列）、研究人才培养（T 系列和 F 系列）、研究职业发展（K 系列）3 种类型，各类型的子类信息及资助金额，见表 1。可以看出，2017—2021

年间，R01、ZIA、T15、R21、K01 这 5 种类型获得的资助数量较多，共 816 项，占总资助数量的 86.35%。这表明 NLM 在注重基础研究和应用研究的同时，也注重对高风险、高回报研究的布局 and 从多个维度对医学信息学、数据科学、计算科学研究人才培养的布局。就资助金额而言，R01、ZIA、T15、R21、K015 类研究项目资助金额数量较多，共获得 36 148.31 万美元支持，占总资助金额的 93.23%。就平均经费额度而言，DP2、R24、ZIA 这 3 类项目较多，其中，校内研究项目平均资助额度为 79.18 万美元，DP2 和 R24 项目数量虽然不多，但平均资助额度分别为 443.25 万美元和 112.89 万美元。这表明 NLM 对创新性强的引领性研究项目和资源建设相关的研究项目有重点布局。

表 1 2017—2021 年 NLM 资助项目类型分布情况

类型	数量（项）	金额（万美元）	平均资助金额（万美元）	类型	数量（项）	金额（万美元）	平均资助金额（万美元）
DP5	3	122.89	40.96	R43	14	314.85	22.49
DP1	2	151.07	75.54	R13	14	27.25	1.95
DP2	1	443.25	443.25	R15	8	262.06	32.76
F31	34	138.46	4.07	R44	7	492.48	70.35
F30	2	10.41	5.20	R42	3	197.75	65.92
K01	40	655.87	16.40	R56	3	109.53	36.51
K99	13	120.28	9.25	R41	3	62.71	20.90
K22	1	1.75	1.75	R25	2	42.04	21.02
R01	436	15 947.81	36.58	R24	1	112.89	112.89
R21	54	967.58	17.92	T15	129	6 145.18	47.64
R00	18	416.16	23.12	ZIA	157	12 431.87	79.18

注：DP5 为面向新拥有博士学位研究人员早期独立研究项目，DP1 为 NIH 院长杰出人才研究项目，DP2 为 NIH 院长创新人才研究项目，F31 为博士研究基金，F30 为硕博研究基金，K01 为研究人才发展项目，K99 为博士后研究基金，K22 为研究职业人才孵化项目，R01 为竞争性研究项目，R21 为探索性/发展性研究项目，R00 为独立研究孵化项目，R43 为小型企业第 1 阶段创新研究资助，R13 为研究性会议资助，R15 为面向教育机构的小规模研究项目，R44 为小型企业第 2 阶段创新研究资助，R42 为小型企业技术转让第 2 阶段研究支持，R56 为高优先度短期研究项目，R41 为小型企业技术转让第 1 阶段研究支持，R25 为教育相关的研究项目，R24 为资源相关的研究项目，T15 为专业研究人 才培训项目，ZIA 为校内研究项目。来源：NIH. Activity Codes Search Results. https://grants.nih.gov/grants/funding/ac_search_results.htm。

3.2.3 资助研究机构 2017—2021 年间，NLM 资助研究机构涵盖医学领域研究机构、高等院校、创新企业及其他非营利性机构等多种类型，其中医学院校、非营利性研究机构、生物医学工程类研究机构、医院 4 类是资助重点，获得 824 项资助，占资助总量的 87%，见图 3。从机构获得的资助数量来

看，美国国立医学图书馆、斯坦福大学、哥伦比亚健康科学学院、宾夕法尼亚大学、匹兹堡大学匹兹堡分校、华盛顿大学、丹佛科罗拉多大学、俄勒冈州健康与科学大学、印第安纳大学－普渡大学印第安纳波利斯分校、密歇根大学安娜堡分校 10 所医学研究机构和大学是获得资助最多的机构，共获得

413 项资助，占资助总量的 43.70%。从机构获得的资助类型来看，Top10 机构中美国国立医学图书馆、斯坦福大学、哥伦比亚大学健康科学学院、宾夕法尼亚大学 4 所研究机构和高等院校获得 20 项以上的研究资助，表明这些院校在医学信息学、数据科学、计算医学等领域拥有较强的研究实力；俄勒冈州健康与科学大学、哥伦比亚健康科学学院、华盛顿大学、丹佛科罗拉多大学 4 所高等院校获得 10 项以上的研究人才培养项目资助，表明这些院校是 NLM 医学信息学领域科研人才培养的重要基地，见表 2。

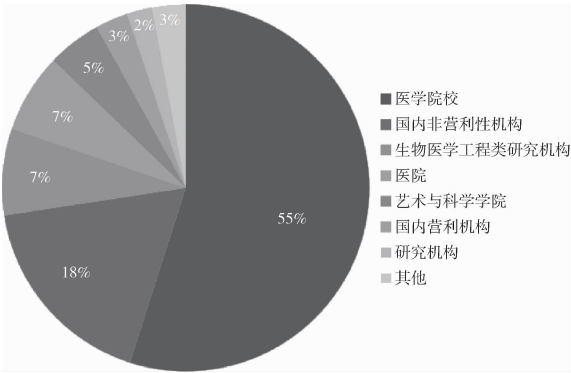


图 3 2017—2021 年 NLM 资助机构类型分布情况

表 2 2017—2021 年 NLM 资助 Top10 机构情况

机构	项目数（个）	项目经费 （万美元）	平均经费额度 （万美元）	研究系列 （个）	研究人才培养系列 （个）
美国国立医学图书馆	156	12 352.38	79.18	156	0
斯坦福大学	39	1 926.59	49.40	30	9
哥伦比亚健康科学学院	37	1 628.74	44.02	24	13
宾夕法尼亚大学	34	1 021.00	30.03	26	8
匹兹堡大学匹兹堡分校	31	1 106.80	35.70	17	14
华盛顿大学	30	959.19	31.97	18	12
丹佛科罗拉多大学	23	943.98	41.04	12	11
俄勒冈州健康与科学大学	22	812.87	36.95	6	16
印第安纳大学－普渡大学印第安纳波利斯分校	21	615.29	29.30	13	8
密歇根大学安娜堡分校	20	403.15	20.16	12	8

3.3 研究资助领域布局

3.3.1 总体布局分析 在 NLM 2017—2021 年的项目资助主题中，网络和信息技术研发、临床研究、生物工程 3 个主题出现次数最多，并且与机器学习和人工智能、数据科学、遗传学、患者安全、健康服务、健康差距、少数人群健康、妇女健康、行为与社会科学、传染疾病等主题分类关联度较

高，见图 4。这表明网络和信息技术研发、临床研究、生物工程 3 个主题是 NLM 近年研究战略布局的重点，而强化卫生服务、缩小健康差距、改善少数族群健康、促进妇女健康、关注覆盖生命全周期的公共卫生挑战、重视数据科学在生物医学研究中的驱动作用、推进机器学习和人工智能在支持科学发现和临床护理中的应用、保护患者安全等多个主题是贯穿 NLM 研究战略布局的重要举措。

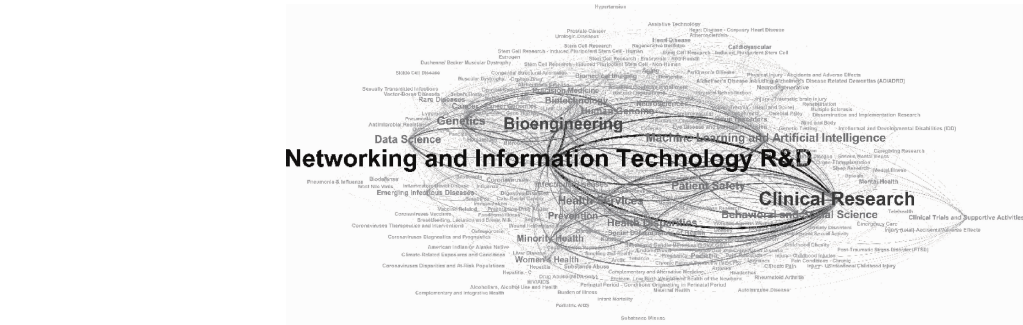


图 4 2017—2021 年 NLM 项目资助主题共现分布

3.3.2 加大网络和信息技术研发 网络和信息技术研发作为核心研究主题的地位凸显,与生物工程、机器学习和人工智能、临床研究、健康服务、数据科学、健康差距等高频主题之间关系紧密,体现 NLM 对生物信息学、计算健康和计算生物学领域先进网络和信息技术研发的布局。如针对人工智能技术应用于健康大数据中存在的系统性偏差、盲点、推理过程及结果的公平性、准确性、安全性等挑战, NLM 实施了增强健康数据发现的计算和统计方法研究专项,重点资助适用于健康大数据的算法或统计方法、纠正偏差或补偿缺失数据的方法、加强整合生物医学和社会/行为数据的统计分析及其他算法等的研发;为了支撑 NLM 2017—2027 年战略规划目标的实现, NLM 推行生物医学研究资产大规模治理计算方法研究专项,进行存储和检索技术、标注方法、数据标准、可视化工具和其他高级数据管理方法的研究资助,加强实现 NLM 基础架构现代化的技术路径研究探索,以及支撑 21 世纪数字图书馆利用的新技术研究。

3.3.3 促进数据驱动的生物医学科学发现 数据科学、生物工程、遗传学、人类基因、生物技术、生物医学图像、癌症、罕见病、精准医学等高频主题在项目中的共现,体现了 NLM 从数据、标准、技术、方法等方面促进生物医学科学发现的研究布局,例如建设生物工程领域有效和可靠的数据资源、研发基于生物医学大数据的知识提取方法和工具、支持计算生物学和信息科学领域应用方法研究等。具体包括:支持基因序列分析、蛋白质结构/功能分析、化学信息学和基因组分析等方面的研究;与美国心肺血液研究所联合资助蛋白质行为研究,致力于通过理解蛋白质形状变化机理来帮助科学家了解癌症和阿尔茨海默症等疾病的分子基础;资助数据库检索算法、进化数学模型、比较基因组学、分类树、群体遗传学和系统生物学等方面的研究。

3.3.4 推进人工智能在生物医学发现、医疗服务中的应用 机器学习和人工智能、神经科学、脑疾病、老龄化、眼病和视觉紊乱、基因检测等高频主题共现,体现了 NLM 利用人工智能重构医疗卫生

服务、优化人工智能技术和方法的研究布局。当前的研究内容包括:将人工智能应用于疾病预防,如通过挖掘医疗健康电子记录和可穿戴设备实时获取的数据等方式来监测健康状况;开发用于早期诊断和干预的人工智能方法;开发新的人工智能方法分析来自学术文献和大规模多源异构数据集中的图像,推进科学发现;利用深度学习算法识别癌症、眼病等疾病医学影像中的病灶,研究深度学习在利用“嘈杂数据”时的局限性;构建生物医学信息分析系统,辅助识别耐药性以降低治疗风险;分析医疗健康电子记录数据,辅助诊断治疗决策,制定合理给药方案,避免不合理的用药问题。

3.3.5 基于健康信息服务的疾病预防和健康促进 网络和信息技术研发、临床研究、健康服务、健康差距、预防、传染性疾病、少数民族健康、妇女健康等高频主题的共现,体现 NLM 利用先进网络和信息技术的提升健康服务水平和缩小健康差距方面的研究布局。作为国家医学图书馆, NLM 通过提供优质健康信息资源、组建国家医学图书馆网络、研发支撑医学科研人员的研究工具、培养生物医学信息学科领域人才等方式来优化健康信息服务,具体包括:缩小少数民族在医疗知识和数据获取方面的差距,促进妇女健康,以医学院校为基础提供生物医学信息学科研究人才专业培训,促进重大突发公共卫生事件中优质健康信息的可及性,实施数据驱动的新型新型冠状病毒肺炎研究专项,进行严重急性呼吸系统综合征冠状病毒的识别、预测、传播及影响的研究。

3.3.6 建设支持临床研究的大数据设施 临床研究、网络和信息技术研发、患者安全、行为和社会科学、临床实验与支持活动等高频主题共现,体现了 NLM 通过将医学大数据资源、医学数据标准、医学数据分析工具等方面的研究成果转化为新的治疗方法、产品和实践,支持临床决策的研究布局。围绕 NIH 战略规划提出的“理解影响人类健康基本过程以促进和重塑健康”的基本目标, NLM 从生物医学信息角度建设覆盖基础科学、疾病预防与健康促进、疾病干预与治疗的大数据基础设施,拉长生物医学数据表征和治理链条,获取、组织、呈现和

提供从细胞到个人再到社会的全生命周期健康数据,建设涵盖基因型和表型数据库、临床实验数据库、生物医学文献数据等生物医学信息资源;资助面向临床大数据的分析方法和工具,开展使用移动技术诊断结核病和疟疾、基于移动终端拍摄的图片识别药品信息、基于 NLM 术语资源和工具的多源去标识处方和患者报告数据分析等方面的研究。

4 启示与借鉴

4.1 立足生物医学发现和数据驱动健康前沿需求

云计算、大数据、人工智能、移动互联网、物联网等新型数智技术在生物医学领域的深入应用,数据科学、开放科学的兴起^[13],给医学图书情报机构的服务和管理带来全新挑战。促进健康医疗大数据在疾病预防、分级诊疗、诊断决策、疾病诊疗、健康管理、临床研究、公共卫生风险预警、病毒溯源等领域的应用,实现医学领域数据、医疗行业数据、医学学科数据、互联网数据和基于智能终端及感知设备的动态数据等多源异构数据的互联互通,支撑数据密集型医学研究和真实世界数据研究的推进^[14-15],成为医学图书情报机构服务赋能的关键所在。医学图书情报机构需要立足生物医学发现和数据驱动健康前沿需求,加强生物医学大数据生态环境建设研究布局,聚合覆盖生命健康全过程的生物医学大数据资源,研发促进生物医学数据创新应用的人工智能、机器学习、计算生物学算法等技术工具,制定多源数据规范、组织、融合标准。

4.2 推行分层次研究资助策略

健康中国战略的实施和新型数智技术的融合从广度和深度两个层面驱动着医学图书情报机构服务的拓展^[16],要求医学图书情报机构的科学研究实施不仅仅局限于促进自身学科建设和发展^[17],还应该面向健康全链条发展需求,实施供给推动型与需求拉动型科研资助新模式,瞄向经济社会发展需求“生产”和“传输”医学信息学新知识。一方面,立足科研活动发展基本规律,坚持可持续发展基本理念,面向医学领域的教育、研究及临床机构,提

供研究孵化增强型资助、应用导向型研究资助、引领性创新型研究资助等多种类型的资助机制,提供支撑图书情报服务能力的基础研究和应用技术研究;另一方面,面向健康全链条发展需求,提供支撑产业化应用、技能训练与经验提升的资助类型,通过与健康产业间的创新及技术转让研究、医学信息学研究人才培养等多种方式,发挥医学图书情报机构对健康全链条的健康知识和人才的辐射作用。

4.3 实施多通路数据驱动型研究人才培育机制

国家卫生健康委员会发布的《“十四五”卫生健康人才发展规划》中指出,促进卫生健康事业高质量发展、推动健康中国建设,人才是关键,并将适应卫生健康行业新技术发展趋势的医学信息技术人才建设作为加强卫生健康人才队伍建设的重要组成部分^[18]。就我国医学信息学人才建设现状而言,虽然我国已经培养了一批支撑医学信息化建设的医学信息人才队伍,但与新一代信息技术和数字化转型的时代需求相比,仍存在供需失衡、人才结构有待优化、人才培养机制需要拓展等方面的问题^[19],NLM 依托馆内和馆外研究项目实施的多通路数据驱动型研究人才培育机制为我国医学图书情报机构进行医学信息学人才培养提供了经验借鉴。

NLM 面向数据驱动健康战略发展目标,致力于打造数据驱动型研究人才多通路培育机制,通过研究职业发展、培训计划、本硕博研究基金等形式,加大对生物医学信息学研究人才培养的力度。如通过馆内研究资助项目培养在医学信息学、数据科学和医学知识管理领域发挥引领作用的专业研究队伍;通过独立研究孵化项目为生物医学信息学研究人员提供研究职业转化资助,为他们独立承担研究项目提供前期基础;推行基于医学院校的生物医学信息学和数据科学培训计划来增强医学信息学研究人才储备基础。截至目前,NLM 已经推出涵盖博士后、研究生、本科生等多层次生物医学信息学和计算生物学研究领域的馆内研究人才培养计划,与加州大学圣地亚哥分校等 16 所高等院校联合设置的生物医学信息学和数据科学研究培训计划。

4.4 打造多元协同研究生态体系

新一轮的科技革命和产业变革将科学研究由学术机构自由探索为主的“小科学”时代，带入到聚合社会各界力量协同创新的“大科学”时代^[20]。聚焦战略任务需求，统筹布局医学信息学研究创新，有效联动政府机构、科研院所、医学院校、医疗卫生机构、医学领域创新企业等创新单元，打造辐射医学创新全链条的医教研系统研究生态体系，是当前 NLM 推行“有组织的科学研究”的重要形式。一方面，加强与政府之间的联系，面向政府优先事项进行研究布局；另一方面，作为 NIH 的信息学、数据科学和知识管理中心，NLM 加强与 NIH 战略需求的对接，围绕 NIH 的关键战略计划“我们所有人”（All of Us）、脑计划、数据科学战略规划等进行研究资源分配，如面向 All of Us 的战略举措，提供缩小少数族群健康和卫生差距的研究项目；再次，通过馆内研究计划和馆外研究计划，整合国内外医学信息学领域优秀的研究机构、医学院校、医学创新企业及其他非营利性机构等多种类型科研实体，共同致力于生物医学信息学领域科学问题的解决。

5 结语

论文基于 NLM 近 5 年发布的科研战略规划、研究专项及资助科研项目等信息，综合运用统计分析、聚类分析等文献计量学方法技术，从研究投入的视角分析 NLM 在数据驱动创新背景下的研究布局特征、动向及可借鉴的经验，以期为我国医学图书馆和医学信息学科领域科研规划决策提供参考。研究发现，自 2017 年以来，NLM 面向实现数据驱动发现和健康的战略目标、NIH 开放数据科学计划和国家优先事项，瞄准生物医学和健康数据科学前沿领域，通过馆内研究计划和馆外研究计划，逐年加大对医学信息学、数据科学、计算科学领域研究项目的资助力度；高度重视研究推进、研究人才培养和研究职业发展等类型科研活动，兼顾基础研究、应用研究和创新研究之间的平衡，重点加强对医学院校、非营利性研究机构、生物医学工程类研

究机构、医院 4 类研究机构的资助；围绕网络和信息技术研发、临床研究、生物工程 3 个关键主题进行研究资源配置，在生物医学大数据治理机制构建、临床决策工具研发、泛在生物医学信息平台建设等方面引领创新。

由于论文分析的侧重点是从研究资源供给和投入视角研判 NLM 研究布局，因此主要选取 NLM 在 2017—2021 年发布的、能公开获取的科研战略规划、研究专项信息及资助的科研项目数据作为分析数据源，后续将进一步拓展分析视角和分析数据源类型，比如将分析时间区间内的 NLM 机构发表的论文、申请的专利、研究成果产生的社会影响等方面的数据纳入分析体系，从投入—产出视角分析 NLM 的研究布局落地成效与产生的社会影响；也可以将分析时间区间内 NLM 研究项目申请数量纳入分析体系，通过分析 NLM 不同年份和主体的申请成功率分布来进一步深化识别 NLM 的科研资源分配情况。在分析方法上，论文对于 NLM 规划文本主要采取人工内容研读的方式获取信息，后续将引入文本分析法以提升分析的客观性和效率。

参考文献

- 1 崔佳伟, 吴思竹, 邬金鸣, 等. 美国国立医学图书馆 2017—2027 年发展战略规划分析与探讨 [J]. 医学信息学杂志, 2019, 40 (3): 75–81.
- 2 郭进京, 陈秀娟, 任慧玲, 等. 开放科学环境下研究型图书馆资源建设与服务分析——以 NLM 为例 [J]. 图书情报工作, 2018, 62 (7): 41–48.
- 3 NLM. A platform for biomedical discovery and data-powered health [EB/OL]. [2022-05-21]. https://www.nlm.nih.gov/pubs/plan/lrp17/NLM_StrategicReport2017-2027.html.
- 4 池慧, 任慧玲, 刘懿. 面向医学科技发展, 做好医学图书馆“十四五”规划 [J]. 数字图书馆论坛, 2021 (5): 31–36.
- 5 NIH. NIH RePORTER [EB/OL]. [2022-04-10]. <https://reporter.nih.gov/>.
- 6 NIH. Activity codes search results [EB/OL]. [2022-04-10]. https://grants.nih.gov/grants/funding/ac_search_results.htm.

(下转第 97 页)

- 31-34.
- 2 陆道坤. 课程思政推行中若干核心问题及解决思路 [J]. 思想理论教育, 2018 (3): 64-69.
- 3 邱伟光. 课程思政的价值意蕴与生成路径 [J]. 思想理论教育, 2017 (7): 10-14.
- 4 张敬普, 董国忠, 薛冰. 思政教育融入《大数据概论》课程教学的探索与实践 [J]. 信息系统工程, 2021 (7): 174-176.
- 5 李大舟, 高健. 课程思政融入数据科学与大数据技术专业实践教学探索 [J]. 中国教育信息化, 2021 (20): 93-96.
- 6 王文乐, 雷刚, 蒋长根, 等. 大数据技术课程群的“导师制”课程思政教学模式探究 [J]. 计算机教育, 2021 (10): 56-59.
- 7 申玉斌, 李曼, 薛宾田. 课程思政融入大数据技术通识课程的教育研究探析 [J]. 教育观察, 2021, 10 (37): 48-50.
- 8 季伟东, 孙小晴, 张军, 等. 大数据方向工程硕士实践教学模式研究 [J]. 计算机教育, 2019 (9): 51-54, 59.
- 9 李辉, 张标. 涉农高校数据科学与大数据技术专业人才培养思考 [J]. 高等工程教育研究, 2019 (5): 16-22.
- 10 冯小洁. 基于大数据思维的大数据技术原理与应用教学改革 [J]. 计算机教育, 2020 (4): 133-137.
- 11 白风伟, 张岩. 基于成果导向的大数据技术与应用专业实验室建设研究 [J]. 高等教育, 2020 (10): 34-35.
- 12 梁艳, 傅美婷, 尹莉. 大数据技术课程开放式教学方式的研究 [J]. 辽宁科技学院学报, 2020 (8): 50-52.
- 13 吴婷, 陈丹, 王明军. 课程思政在大数据技术与应用课程中的融合方法探究 [J]. 产业与科技论坛, 2020, 19 (17): 91-92.
- 14 欧阳林艳, 黄风华. 新工科背景下数据科学与大数据技术专业应用型人才培养的思考与改革 [J]. 信息系统工程, 2020 (12): 163-164.
- 15 朱常鹏, 刘元超, 李刚. Spark 实时大数据处理技术在线开放课程建设与实践 [J]. 现代信息科技, 2021, 5 (12): 195-198.
- 16 林刚, 陈群, 曾志平. 应用型本科院校大数据技术课程思政建设 [J]. 计算机教育, 2021 (9): 32-36.
- 17 陆悠, 傅启明, 邹恩岑. 多课程联动的大数据技术课程实践教学方法研究 [J]. 计算机教育, 2019 (6): 53-57.

(上接第 86 页)

- 7 NLM. NLM congressional justifications [EB/OL]. [2022-05-25]. <https://www.nlm.nih.gov/about/appropriations.html>.
- 8 NLM. Grants and funding; extramural programs (EP) [EB/OL]. [2022-06-20]. <https://www.nlm.nih.gov/ep/ep-overview.html>.
- 9 NLM. NLM extramural programs (EP) [EB/OL]. [2022-06-02]. <https://www.nlm.nih.gov/ep/index.html>.
- 10 National Institutes of Health, National Center for Health Statistics. Estimates of funding for various research, condition, and disease categories (RCDC) [EB/OL]. [2022-05-26]. <https://report.nih.gov/funding/categorical-spend-ing#/>.
- 11 NLM. NCBI computational biology branch [EB/OL]. [2022-05-26]. <https://www.ncbi.nlm.nih.gov/research/>.
- 12 NLM. About LHCBC [EB/OL]. [2022-06-10]. <https://lhncbc.nlm.nih.gov/LHC-about/about.html>.
- 13 赵栋祥. 《美国国家医学图书馆战略规划: 2017—2027》的解读与启示 [J]. 图书情报工作, 2019, 63 (5): 138-146.
- 14 戴明锋, 孟群. 医疗健康大数据挖掘和分析面临的机遇与挑战 [J]. 中国卫生信息管理杂志, 2017, 14 (2): 126-130.
- 15 钱庆. “健康医疗大数据管理与应用”专题序 [J]. 数据分析与知识发现, 2020, 4 (12): 1.
- 16 陈锐, 冯占英, 胡畔畔, 等. “健康中国”战略中医学图书馆的挑战和机遇 [J]. 中华医学图书情报杂志, 2017, 26 (6): 1-6.
- 17 吕茜倩, 张沁兰, 易雪媛. “健康中国”战略下医学院校图书馆辐射作用研究 [J]. 医学信息学杂志, 2020, 41 (2): 76-80.
- 18 国家卫生健康委员会. “十四五”卫生健康人才发展规划 [EB/OL]. [2022-08-24]. <http://www.nhc.gov.cn/renshi/s7753/202208/b208219087084501a49110e6f01d22c2.shtml>.
- 19 刘辉. 我国高层次医学信息人才培养体系建构的思考 [J]. 医学信息学杂志, 2021, 42 (11): 2-6.
- 20 万劲波, 张凤, 潘教峰. 开展“有组织的基础研究”: 任务布局与战略科技力量 [J]. 中国科学院院刊, 2021, 36 (12): 1404-1412.