

# 国外科学数据管理与共享政策及对我国精准医学数据管理的启示\*

孙小康 吴思竹 修晓蕾 崔佳伟 钱庆

(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〕** 选取美、英两国生物医学科研资助机构的部分数据管理与共享政策, 对其基本特征及政策概况、数据汇交、数据保存与管理、数据共享以及数据安全与保护 5 大类、25 个具体指标进行详细对比分析, 结合我国精准医学专项基本情况, 获得政策制定方面的启示, 以期为我国精准医学专项数据政策制定提供参考。

**〔关键词〕** 科学数据; 数据管理; 数据共享; 数据政策; 精准医学

**〔中图分类号〕** R-056 **〔文献标识码〕** A **〔DOI〕** 10.3969/j.issn.1673-6036.2018.04.011

**Overseas Scientific Data Management and Sharing Policies and Its Enlightenment to Precision Medical Data Management in China** SUN Xiao-kang, WU Si-zhu, XIU Xiao-lei, CUI Jia-wei, QIAN Qing, Institute of Medical Information, Chinese Academy of Medical Sciences/Peking Union Medical College, Beijing 100020, China

**〔Abstract〕** Selecting some data management and sharing policies from biomedical scientific study funding agencies in both USA and Britain, the paper carries out detailed comparative analysis of its basic characteristics and the 25 specific indexes of the 5 categories, including policy overview, intersection point of data, data storage and management, data sharing and data security and protection. Combining with the specific fundamental situation of precision medicine in China, it gains enlightenment in policy making to provide references for the formulation of specific data policy of precision medicine in China.

**〔Keywords〕** Scientific data; Data management; Data sharing; Data policy; Precision medicine

**〔修回日期〕** 2018-01-08

**〔作者简介〕** 孙小康, 硕士研究生; 通讯作者: 钱庆, 研究员, 发表论文 60 余篇。

**〔基金项目〕** 国家重点研发计划精准医学研究重点专项“重大疾病精准医学数据库群”课题(项目编号: 2016YFC0901602); 北京协和医学院协和青年基金“面向精准医学的数据汇交与共享政策研究”(项目编号: 2017330018)。

## 1 引言

2015 年年初奥巴马在国情咨文中从国家战略层面提出“精准医学计划”(Precision Medicine Initiative, PMI), 引发世界各国的广泛关注和热议。精准医学作为医学科技发展的前沿, 不仅是公众健康的需求和临床发展的要求, 更是医学自身发展的客观必然, 已成为世界各国新一轮科技竞争的战略制高点<sup>[1-2]</sup>。我国及时把握科技前沿, 部署中国版的精准医学计划, 将其纳入国家“十三五”规划并在

2016 及 2017 年度国家重点研发计划中均将“精准医学研究”列为重点专项之一<sup>[3]</sup>。

随着我国精准医学专项的实施，将产生由不同技术和方法获取的、被不同领域科学家收集的、停留在不同理论和信息层面的数据<sup>[4]</sup>。这些数据既是各专项队列重要的成果产出，也是日后开展医学研究工作宝贵的知识来源，如不能及时汇交并且合理共享，将难以发挥其最大价值，造成人力、物力、财力的极大浪费。精准医学数据汇交与共享是一项复杂的活动，不能单靠技术的进步而实现，还需要运用政策的强制手段进行保障和规范。但我国至今尚未发布适用于精准医学研究专项数据汇交与共享的政策性文件，势必影响专项成果的管理与转化。基于此本研究将对国外生物医学科研资助机构数据政策进行深入调研，详细剖析政策结构和内容，分

析我国精准医学数据汇交与共享的实际情况，获得政策制定方面的部分启示，以期为我国精准医学专项数据政策的制定提供参考。

## 2 调研结果

### 2.1 政策文本选择

本研究选取美、英两国的生物医学科研资助机构数据管理与共享政策作为调研对象，其原因是美、英两国在科研数据管理与共享的实践上起步较早，已建立完备的政策条例，且均开展精准医学相关研究项目。通过网络调研，根据易获取性和政策内容的详尽程度，最终选取 10 个政策文本作为深入分析的对象，生物医学科研资助机构数据管理与共享政策文本，见表 1。

表 1 美、英两国生物医学科研资助机构数据管理与共享政策文本

国家	编码	政策名称	颁布机构	制定/修订时间
美国	P01	NIH 数据共享政策及实施指南 <sup>[5]</sup>	美国国立卫生研究院 (NIH)	2003 - 03 - 05
	P02	全基因组关联研究数据共享政策 <sup>[6]</sup>		2007 - 08 - 28
	P03	NIH 基因组数据共享政策 <sup>[7]</sup>		2014 - 08 - 27
	P04	CDC/ATSDR 数据发布与共享政策 <sup>[8]</sup>	疾病预防控制中心 (CDC)	2005 - 09 - 07
	P05	NIDDK 数据共享政策 <sup>[9]</sup>	国立糖尿病、消化及肾脏疾病研究所 (NIDDK)	2013 - 07
	P06	关于完整申报材料所需数据管理计划的更新信息 <sup>[10]</sup>	国家科学基金会 (NSF) 生物科学部	2015 - 10 - 01
英国	P07	CRUK 数据共享指南 <sup>[11]</sup>	英国癌症研究中心 (CRUK)	-
	P08	MRC 数据共享政策 <sup>[12]</sup>	医学研究理事会 (MRC)	2016 - 09
	P09	BBSRC 数据共享政策 <sup>[13]</sup>	英国生物技术与生物科学研究理事会 (BBSRC)	2017 - 03
	P10	数据、软件及材料管理与共享政策 <sup>[14]</sup>	惠康基金会	2017 - 07 - 10

### 2.2 政策文本侧重点

所调研的政策虽然均是对数据管理及共享进行规定，但侧重点有所差异。根据各政策文本侧重点的不同将其划分为：对数据管理与共享的声明、对数据管理与共享多方面的规定以及对数据管理与共享特定方面的规定 3 种类型。各政策文本侧重点，见表 2。

表 2 各政策文本侧重点

侧重的方面	政策名称
对数据管理与共享的声明	MRC 数据共享政策 (偏向于对共享原则的声明)

续表 2

对数据管理与共享多方面的规定	数据、软件及材料管理与共享政策 NIH 数据共享政策及实施指南 全基因组关联研究数据共享政策 NIH 基因组数据共享政策 CDC/ATSDR 数据发布与共享政策 CRUK 数据共享指南 BBSRC 数据共享政策
对数据管理与共享特定方面的规定	NIDDK 数据共享政策 (偏向于对共享时间的规定) 关于完整申报材料所需数据管理计划的更新信息 (偏向于对数据管理计划的规定)

### 2.3 政策内容分析

2.3.1 框架 参照 Dietrich D 等人<sup>[15]</sup>、陈大庆<sup>[16]</sup>以及魏悦等人<sup>[17]</sup>关于科研资助机构数据管理与共享政策的分析框架，兼顾精准医学专项数据生命周期及数据汇交与共享的实际需求，最终选取政策概况、数据汇交、数据保存与管理、数据共享以及数据安全与保护 5 大类、25 个具体指标作为本研究的分析框架。政策内容分析框架，见图 1。

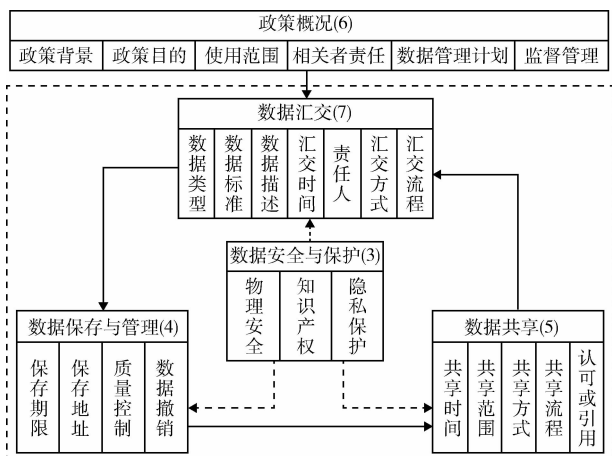


图 1 政策内容分析框架

2.3.2 结果 运用上述政策内容分析框架，通过

研读政策文本，提炼其中的内容要点。美、英生物医学科研资助机构数据管理与共享政策主要内容，见表 3。由表中的汇总数据可以看出，内容覆盖率超过 50% 的政策只有 4 项，而内容覆盖率最低的两项政策分别是：《关于完整申报材料所需数据管理计划的更新信息》(12%) 和《NIDDK 数据共享政策》(20%)。究其原因，可能是这两项政策的颁布机构均为二级资助机构，如 NIDDK 隶属于 NIH，而 NSF 生物科学部只是 NSF 的一个分支，因此沿袭上层机构政策的相关规定，而在自身政策制定时更多地考虑需要着力强调或保障的方面。其从具体政策内容方面看，政策覆盖率最高的前 5 项政策内容依次为：政策目的、数据管理计划、监督管理、数据共享时间以及隐私保护，其覆盖率均为 80%，说明这 5 方面的内容在数据管理共享过程中更受关注，需要政策加以规定和引导。而覆盖率最低的前 5 项政策内容为：汇交责任人 (0)、汇交流程 (0)、共享流程 (0)、汇交方式 (10%) 和数据物理安全保证 (10%)，其原因可能是该方面的内容比较具体，资助机构层面的政策少有涉及。此外还可以看出政策概况性内容覆盖率相较其他部分要高一些，也从一定程度上反映出所调研的这些政策内容更偏向宏观层面的描述，对于细节的规定尚显不足。

表 3 美、英两国生物医学科研资助机构数据管理与共享政策主要内容

政策编码	政策概况						数据汇交					数据保存与管理				数据共享				数据安全与隐私				汇总 (%)		
	政策背景	政策目的	适用范围	相关者责任	数据管理计划	监督管理	数据类型	数据标准	数据描述	汇交时间	汇交责任人	汇交流程	汇交方式	保存期限	保存地址	质量控制	数据撤销	共享时间	共享范围	共享方式	共享流程	引用或致谢	数据物理安全		知识产权	隐私保护
P01	-	√	√	-	√	√	√	-	√	-	-	-	√	-	-	-	√	√	√	-	-	-	-	√	44	
P02	√	√	√	√	-	√	√	√	√	√	-	-	-	√	√	√	√	√	-	-	√	√	√	√	72	
P03	-	√	√	-	√	√	√	-	√	√	-	-	-	√	-	√	√	-	√	-	√	-	√	√	60	
P04	√	√	√	√	√	√	-	√	√	-	-	-	-	√	√	-	-	-	√	-	-	-	√	√	52	
P05	-	√	-	-	√	-	-	-	-	-	√	-	-	√	-	-	√	-	-	-	-	-	-	-	20	
P06	√	-	-	-	√	√	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	12	
P07	-	√	√	-	√	-	-	√	√	-	-	-	√	-	-	-	√	-	√	-	-	-	√	√	40	
P08	√	√	√	-	√	√	-	-	-	-	-	-	-	√	-	√	-	-	-	-	-	-	√	√	36	
P09	√	-	-	-	√	√	√	√	√	-	-	-	√	√	√	-	√	-	√	-	√	-	√	√	56	
P10	√	√	-	-	√	√	-	-	-	-	-	-	√	-	-	√	-	-	-	-	√	-	√	√	36	
汇总 (%)	60	80	60	20	80	80	50	40	60	20	0	10	0	30	60	40	20	80	20	50	0	40	10	70	80	-

2.3.3 政策概况 (1) 政策背景。一般位于政策开头,用以介绍与政策或颁布机构相关的一些信息。如在《全基因组关联研究数据共享政策》的背景部分对全基因组关联研究、政策制定过程以及政策所要重点解决的问题等内容进行了介绍;而《数据、软件及材料管理与共享政策》则侧重于对惠康基金会的介绍。(2) 政策目的。调研政策中关于政策目的的描述大致可以分为 3 类:①平衡利益相关者的权益,为数据收集者提供一个合理的共享时滞期(如 NIDDK 的政策)。②促进利益最大化,加快研究成果向知识、产品和改善人类健康的应用方面转换(如 NIH 颁布的政策)。③在广泛、自由的数据共享的基础上,保护知识产权以及受试者的隐私(如 CRUK 的政策)。(3) 适用范围。指该政策所要覆盖的研究或项目的集合。一般而言,资助机构的政策适用于其资助的所有研究(如 MRC 的政策),但一些政策在此基础上会有额外的说明或限定。如《NIH 数据共享政策及实施指南》还包括:①为研究目的共享最终研究数据。②NIH 支持的基础研究、临床研究、调查和其他类型的研究。③通过资助、合作协议或合同所获得的任意一年直接经费大于 50 万的研究项目。④2003 年 10 月 1 日以后提交的资助申请。(4) 相关者责任。数据管理与共享过程中涉及的利益相关者包括:数据提供者、科研资助机构、数据使用者以及与之相关的承包商和合作伙伴等。《全基因组关联研究数据共享政策》中对数据提供者和数据使用者的责任进行了详细规定,对其数据的合理共享和规范、安全使用提供了必要保障。(5) 数据管理计划。所调研的政策中除《全基因组关联研究数据共享政策》和《数据、软件及材料管理与共享政策》外,其他政策均对数据管理计划做了说明。如《CRUK 数据共享指南》建议在数据管理计划中重点考虑以下内容:①最终数据集的容量、类型、内容和格式。②数据收集和管理的标准。③解释性文档或辅助材料。④数据共享方法。⑤数据共享的时间。⑥数据长期保存计划。⑦限制数据共享的原因。⑧知识产权及专有数据的说明。⑨隐私、道德以及知情同意方面的考虑等。(6) 监督管理。虽然绝大多数政策涵盖这方面的内

容,但未做详细说明。总结调研政策中监管的主体包含:①资助机构本身。②资助机构内的特定组织,如 NIH 的机构中心、数据访问委员会、NSF 生物科学部项目主管等;监管方式通常是对年度进度报告及最终项目报告进行审查;而对违规行为的具体处罚措施,以上政策均未做出具体说明。

2.3.4 数据汇交 (1) 数据类型。5 项政策中对用于汇交或共享的数据类型进行说明,且数据类型因相应的研究类型而不同。如《NIH 基因组数据共享政策》主要适用于组学领域,因此其数据类型包括:全基因组关联研究、单核苷酸多态性阵列和基因组序列、转录组学、元基因组学、表观基因组学和基因表达数据以及相关关联数据等;而惠康基金会由于资助的研究类型广泛,因此政策中规定的数据类型为发明、数据集、软件和材料以及出版物等。(2) 数据标准。在 4 项包含数据标准描述的政策中,除 CDC 会根据数据归档最佳实践制定专门的数据标准外,其他 3 项均表明建议使用现有的标准。(3) 数据描述。详尽的数据描述能确保他人可以正确地使用数据集,防止滥用、误解和混淆。数据描述信息可分为两大类:①对研究项目的描述(如《全基因组关联研究数据共享政策》)。②对数据采集方式、采用标准、附加文档、各变量定义、分析工具等相关内容的说明。(4) 汇交时间。调研政策中对数据汇交时间的界定甚少,虽然有两项政策包含相关描述,但均未给定具体的时间节点,使用“尽快”、“及时”等模糊性的词汇来定义。(5) 汇交责任人。所有政策中均未来对汇交责任人进行规定,但笔者认为责任到人既便于研究团队内部进行协调组织,提高数据汇交的及时性及数据的质量,也便于监管部门对违规行为进行追责,强化监管力度。(6) 汇交方式。仅 NIDDK 的政策中有关于汇交方式的描述,其政策中指出“对于观察性研究和长期临床试验而言,这种分阶段的数据发布方法尤其重要”。而精准医学专项实施周期长,且数据分批采集,因此这种分阶段汇交的方式同样适用。(7) 汇交流程。调研政策中没有关于数据汇交流程的描述,详细的汇交流程说明可以帮助汇交责任人明晰数据汇交的各个环节,缩短汇交时间,从

而尽早实现数据共享。

2.3.5 数据保存与管理 (1) 保存期限。数据的保存期限与数据的价值、收集难度以及可复制性等多方面因素有关。《NIH 数据共享政策及实施指南》中规定数据在项目结束后最少保存 3 年;《CRUK 数据共享指南》对保存时间的规定是至少 5 年;而《BBSRC 数据共享政策》中要求数据保存期限至少为 10 年。(2) 保存地址。在数据保存地址的选择上,资助机构也有不同规定,大致可以分为两类:①自建数据仓储中心,如 NIH 建立的 GWAS 数据中心和 CDC 建立的信息中心。②在公认的数据库中进行存储,如惠康基金会的规定。(3) 质量控制。对数据进行质量控制能提高数据的可用性,最大程度地发挥数据的价值。调研政策中关于数据质量的描述可以概括为两方面:①CDC 规定要对数据生命周期中的每个阶段进行质量评估,以保证高质量的数据。②MRC 强调除对数据生命周期进行审查以外,还应提供高质量的元数据来增强数据的可用性。(4) 数据撤销。NIH 的两项政策中均有关于数据撤销方面的规定。具体的描述为“如果研究受试者要求撤回知情同意,可从 NIH 指定的数据存储库中删除个人参与者的数据。但是,一些已经用于共享的数据无法撤回”。数据撤销反映了资助机构对研究受试者意愿的高度尊重。

2.3.6 数据共享 (1) 共享时间。对于数据共享时间的描述,大致可以分为 3 类:①以行业内的最佳实践为参考,如 BBSRC 建议数据的共享时间遵循行业的最佳实践,如果不存在最佳实践,则共享时间不晚于 3 年。②不晚于第 1 份研究出版物被接收的时间,如《NIH 数据共享政策及实施指南》的规定。③规定具体的时间段,如数据提交后 6 个月、12 个月、2 年或者 3 年内实施共享。此外,如果是专有数据或者需要申请专利,可以根据政策要求申请延期共享。(2) 共享范围。对于数据共享范围的界定也有所不同,《NIH 数据共享政策及实施指南》中的共享范围是指用于共享的数据的范围,需要依照数据的敏感性和机密性来决定;而《全基因组关联研究数据共享政策》中则是指共享人群的范围,会尽可能广泛地向科学研究界提供,但需要

通过数据访问委员会的审批。(3) 共享方式。《NIH 基因组数据共享政策》及 CDC 的政策中将共享方式分为两种,即公开访问和受限访问;而《NIH 数据共享政策及实施指南》和 CRUK 政策的规定则更加详细分为 4 种:①在首席研究员的主持下进行共享。②通过公共数据仓储共享。③通过数据飞地提供了一个受控的安全环境进行受限共享。④根据数据特点,通过组合的方式进行共享。(4) 共享流程。对于数据共享的流程,所有的政策均没有详细的说明。《全基因组关联研究数据共享政策》虽然有“用户访问数据时需要提交含有数据使用证书的数据访问请求,并交由数据访问委员会审批”的描述,但并未清晰地描述整个共享环节,使用户的访问请求无法得到规范指导,可能会影响数据共享的进程。(5) 引用或致谢。对数据提供者的贡献进行认可是保证数据共享活动持续推进的前提。认可的形式可以为:①口头或书面形式进行说明或致谢(如 NIH)。②对数据来源进行引用(如惠康基金会)。③采用联合署名的方式(如 BBSRC)。

2.3.7 数据安全与保护 (1) 数据物理安全。在数据安全方面,大多数政策更偏向于对隐私和数据机密性方面的介绍,而忽视了物理安全的描述。营造一个安全可靠的物理环境,对于数据的保存和管理至关重要,《全基因组关联研究数据共享政策》中指出“要采用多层次的数据安全措施,如防火墙和独立网络,以保证信息能通过安全传输过程提交给 NCBI 内的高安全网络”。(2) 知识产权。内容涉及数据所有权和专利申请两方面。对于数据所有权的的规定, CDC 和 MRC 均认为公共资助产生的数据应作为公共财富。而在专利申请方面,美国和英国有一些差异,美国鼓励对新技术申请专利,但不希望对数据或者直接由数据中得出的结论申请专利(NIH 政策);英国则鼓励申请专利,但并没有界定哪些内容可以纳入专利的范畴。(3) 隐私保护。对受试者隐私的保护是所有数据共享活动的前提,绝大多数政策均有相关内容的规定。美、英两国也出台相关的法律法规进行保障,如《人类保护条例》、《隐私法》、《HIPPA 法案》以及《公共卫生服务法》等。对隐私的保护贯穿于整个数据生命周期

中,在数据收集前需要与受试者充分沟通,征得其数据使用授权并签署知情。数据提交前要进行充分的匿名化处理,保证能识别身份的直接标识符和间接标识符均已从数据集中剥离;而在数据共享过程中要通过受限访问以及签署数据共享协议等方式,保证数据传播过程中患者隐私的安全性。

### 3 我国精准医学专项数据管理与共享

#### 3.1 精准医学专项基本情况

2016 年科技部启动“精准医学”研究重点专项,由国家卫生计生委医药卫生科技发展研究中心负责该专项的申报与管理工作。截至目前 2016 及 2017 年度申报工作均已结束,两年的专项申报指南<sup>[18-19]</sup>中均设立 5 项主要任务,即新一代临床用生命组学技术研发,大规模人群队列研究,精准医学大数据的资源整合、存储、利用与共享平台建设,疾病防治方案的精准化研究,精准医学集成应用示范体系建设。其中明确提出需要汇交并共享数据的研究包括:自然人群队列研究(8 项)、重大疾病专病队列研究(14 项)、罕见病临床队列研究(1 项)、中国人群参比数据库建设与系统分析(1 项)以及重大疾

病临床样本生命组学数据库(1 项)。

#### 3.2 精准医学专项数据汇交与共享需求分析

通过研读 2016 及 2017 年度精准医学专项申报指南,对其中涉及数据汇交与共享的内容进行汇总。2016 及 2017 年度精准医学专项数据汇交与共享需求,见表 4。数据汇交与共享是精准医学专项的强制要求,专项中明确规定的项目,均必须将数据提交至专项所建立的精准医学大数据平台统一进行管理。申报单位和个人必须签署具有法律约束力的协议予以保证。其中所涉及的数据类型复杂,大致包括问卷调查、体格检查、生物样本、生命组学、临床诊疗、随访数据、监测数据等。而对于相关标准规范,除自然人群队列明确提出有示范对列自行建立标准,其他队列进行参照外,其他项目均没有明确指定。数据的最终应用除支撑相关研究外还会建立一些参比库或信息库,以备后续之用。虽然申报指南中明确要求数据汇交与共享,但说法较为笼统,缺乏对具体实施细节的规定,导致数据在汇交、管理及共享方面无章可循。因此及时制定符合实际需求的精准医学数据汇交与共享政策成为重要的工作。

表 4 2016 及 2017 年度我国精准医学专项数据汇交与共享需求

项目类型	支持数量	标准规范	数据应用
自然人群健康队列	16 年: 3 项 17 年: 5 项	示范队列自行制定,其他参照示范队列	支撑国家科研项目;与专项中各方向资源和数据共享
重大疾病专病队列	16 年: 6 项 17 年: 8 项	统一标准和规范 同上	同上
罕见病临床队列	16 年: 1 项	同上	同上
中国人群参比数据库建设与系统分析	17 年: 1 项	无说明	建立中国人群多组学参比数据库
重大疾病临床样本生命组学数据库	17 年: 1 项	标准化	建立国家级、标准化、可共享和第 3 方的生物样本多层次生命组学信息库

### 4 启示

#### 4.1 政策定位

“精准医学”作为国家重点研发计划,由国家进行经费投入和监管,因此在制定政策时应立足于

国家,保证政策的权威性。但相较于美、英两国我国在科学数据管理与共享方面尚未形成成熟的层级体系,政策内容不能进行相互补充。因此在政策制定过程中需要借鉴多方面的政策内容,不断加以完善,实现内容的横向扩展,使所制定的政策不仅在宏观上具有指导作用,在具体细节上也能进行有效

规范,从而提高精准医学数据管理工作的执行效率。

## 4.2 制定过程

精准医学数据在汇交与共享过程中涉及的利益相关者众多,在政策制定过程中只有充分权衡各相关者权益,才能保证政策被广泛接受。在整个过程中主要的利益相关者为:(1)项目资助方:科技部(原国家卫生计生委医药卫生科技发展研究中心负责“精准医学专项”的申报和管理工作)。(2)数据管理方:“精准医学大数据平台”的管理者。(3)数据汇交方:精准医学各人群研究队列。(4)数据生产方:各队列的入组受试者。(5)数据需求方。这些利益相关者均是政策的最终覆盖者,因此可以参考《全基因组关联研究数据共享政策》政策制定的过程,广泛征集这些利益相关者的意见,尽可能满足绝大多数利益相关者的合理需求,这样才能使政策充分满足我国精准医学专项的实际需求,保证可执行性。

## 4.3 具体内容

4.3.1 政策概况 其内容有助于人们对于政策整体情况以及实施要求的理解。所调研的政策中绝大多数对于该部分的内容均进行了说明。在制定精准医学数据政策时应立足于我国精准医学专项实际情况,以专项申报指南的内容为基础进行规定。为保证政策的落地实施以及适用性,还需要在监管以及相关责任者方面进行详细规定,明确项目资助方、数据管理方、汇交方、生产方以及需求方等主要利益相关者的权利和义务,设立相关部门,以项目组提交的数据管理计划、年度进展报告等文件作为依据,专门负责项目进展的监督以及违规行为的处罚等,监督各方完成其应该履行的职责。

4.3.2 数据汇交 应责任到人,规定各项目组安排专人进行数据的汇交工作,对需要汇交的数据类型、方式、时间以及具体流程等方面进行详细说明。就我国精准医学专项而言,各队列专项实施时间与“精准医学大数据平台”建设同期进行,为保证各项目均顺利进行数据汇交可采用分批汇交的方

式,每批次数据完成收集、审核以及整理等工作后1个月内通过专网进行汇交。此外在数据汇交过程中,另一个重点是数据标准,对于自然人群队列而言,其标准可以参考自然人群示范队列,但专病和罕见病则需要自行构建。为保证各队列的数据标准与“精准医学大数据平台”相适应,在制定标准的过程中可以采取多方合作的方式,共同参与标准的制定决策,以保证标准的一致性。

4.3.3 数据保存与管理 精准医学数据具有极高的研究价值,且获取不易,因此可以在精准医学大数据平台中进行永久保存。为保证数据质量,除为数据提供高质量的元数据外,还要在数据生命周期的每个阶段对其进行质量审核。在数据保存期间,任何人不得随意删除平台中的任何数据,除非受试者本人提出申请,经过项目资助机构以及数据管理平台的审批后方可删除该受试者的个人记录。

4.3.4 数据共享 实现数据共享是数据价值挖掘的必经阶段,也是精准医学专项指南中明确规定的內容。由于精准医学数据中蕴含着大量的个人健康信息,因此其共享必须按照严格的流程进行。在共享时间方面,应在各项目组主要成果发表后开始共享,以保护项目组的科研产出。为保证数据不被滥用,需要通过受控的方式仅向科研人员进行共享,而使用数据的科研人员则需要通过数据申请、资格审批、签署数据使用协议等过程才能获得所需数据。在使用数据所产生的成果中也需要通过数据引用、致谢或者联合署名的方式对数据来源进行说明。

4.3.5 数据安全与保护 贯穿于数据生命周期的每一个阶段,在政策的制定过程中需要对其进行详细的规定。从数据收集开始,为数据营造一个安全的物理环境,保证数据在收集、传输、存储以及分析等过程中的安全,定期进行数据备份,做好减灾容灾措施。而对于数据中潜在的隐私信息也要加强保护,即使是脱敏数据,也要防范与其他第3方数据库进行关联而泄露隐私的风险。在数据保护中另一个重要的部分是知识产权的保护,政策中应明确规定数据的所有权以及利用数据产生成果的专利权等内容。精准医学数据是国家经费投入的重要产

出,因此其数据所有权当属于国家,同时应该鼓励使用数据产生的新技术、新方法的专利申请活动,以提高科研人员研究的积极性。

## 参考文献

- 1 中国科学报. 精准医学:我国医学发展的历史机遇 [EB/OL]. [2017-11-20]. <http://www.biotech.org.cn/information/138770>.
- 2 人民网. 精准医学离我们还有多远? [EB/OL]. [2017-11-20]. <http://health.people.com.cn/n1/2016/0829/c21471-28672927.html>.
- 3 仪器信息网. 精准医疗、生命科学正式纳入“十三五”规划 [EB/OL]. [2017-04-06]. <http://www.instrument.com.cn/news/20160321/186765.shtml>.
- 4 钱庆. 精准医学相关的数据管理与知识服务 [J]. 情报工程, 2015, 1 (6): 17-22.
- 5 NIH. NIH Data Sharing Policy and Implementation Guidance [EB/OL]. [2017-11-25]. [https://grants.nih.gov/grants/policy/data\\_sharing/data\\_sharing\\_guidance.htm](https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm).
- 6 NIH. Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWAS) [EB/OL]. [2017-11-25]. <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-07-088.html>.
- 7 NIH. National Institutes of Health Genomic Data Sharing Policy [EB/OL]. [2017-11-25]. [https://osp.od.nih.gov/wp-content/uploads/NIH\\_GDS\\_Policy.pdf](https://osp.od.nih.gov/wp-content/uploads/NIH_GDS_Policy.pdf).
- 8 CDC. CDC/ATSDR Policy on Releasing and Sharing Data [EB/OL]. [2017-11-25]. <https://www.cdc.gov/maso/policy/releasingdata.pdf>.
- 9 NIDDK. NIDDK Data Sharing Policy [EB/OL]. [2017-11-25]. <https://www.niddkrepository.org/static/NIDDK%20data%20sharing%20policy%202013.docx>.
- 10 NSF Biological Sciences. UPDATED Information about the Data Management Plan Required for Full Proposals [EB/OL]. [2017-11-25]. [https://www.nsf.gov/bio/pubs/BIODMP\\_Guidance.pdf](https://www.nsf.gov/bio/pubs/BIODMP_Guidance.pdf).
- 11 CRUK. Data sharing guidelines [EB/OL]. [2017-11-25]. [http://www.cancerresearchuk.org/funding-for-researchers/applying-for-funding/policies-that-affect-your-grant/submission-of-a-data-sharing-and-preservation-strategy/data-sharing-guidelines?\\_ga=2.136838896.1356696494.1511183071-918465613.1511183071](http://www.cancerresearchuk.org/funding-for-researchers/applying-for-funding/policies-that-affect-your-grant/submission-of-a-data-sharing-and-preservation-strategy/data-sharing-guidelines?_ga=2.136838896.1356696494.1511183071-918465613.1511183071).
- 12 MRC. Data Sharing Policy [EB/OL]. [2017-11-25]. <https://www.mrc.ac.uk/documents/pdf/mrc-data-sharing-policy/>.
- 13 BBSRC. BBSRC Data Sharing Policy [EB/OL]. [2017-11-25]. <https://www.bbsrc.ac.uk/documents/data-sharing-policy-pdf/>.
- 14 Wellcome Trust. Policy on Data, Software and Materials Management and Sharing [EB/OL]. [2017-11-25]. <https://wellcome.ac.uk/funding/managing-grant/policy-data-software-materials-management-and-sharing>.
- 15 Dietrich D, Adamus T, Miner A, et al. De-Mystifying the Data Management Requirements of Research Funders [EB/OL]. [2017-11-25]. <http://www.istl.org/12-summer/refereed/>.
- 16 陈大庆. 英国科研资助机构的数据管理与共享政策调查及启示 [J]. 图书情报工作, 2013, 57 (8): 5-11.
- 17 魏悦, 刘桂锋. 基于数据生命周期的国外高校科学数据管理与共享政策分析 [J]. 情报杂志, 2017, 36 (5): 153-158.
- 18 科技部. “精准医学研究”重点专项 2016 年度项目申报指南 [EB/OL]. [2017-11-25]. [http://www.most.gov.cn/tzgt/201603/t20160308\\_124542.htm](http://www.most.gov.cn/tzgt/201603/t20160308_124542.htm).
- 19 科技部. “精准医学研究”重点专项 2017 年度项目申报指南 [EB/OL]. [2017-11-25]. <http://www.most.gov.cn/mostinfo/xinxifenlei/fgzc/gfxwj/gfxwj2016/201610/W020161012586568282225.pdf>.