



美国国立 卫生研究院 (NIH) 数据共享政策

顾立平等 编



科学技术文献出版社
SCIENTIFIC AND TECHNICAL DOCUMENTATION PRESS



美国国立卫生研究院 (NIH) 数据共享政策

>>> 顾立平等 编



科学技术文献出版社
SCIENTIFIC AND TECHNICAL DOCUMENTATION PRESS

· 北京 ·

图书在版编目(CIP)数据

美国国立卫生研究院(NIH)数据共享政策 / 顾立平等编. —北京: 科学技术文献出版社, 2017. 12

ISBN 978-7-5189-3359-4

I. ①美… II. ①顾… III. ①数据共享—信息政策—研究—美国
IV. ①G203

中国版本图书馆 CIP 数据核字 (2017) 第 234263 号

美国国立卫生研究院 (NIH) 数据共享政策

策划编辑: 崔灵菲 责任编辑: 王瑞瑞 责任校对: 文浩 责任出版: 张志平

出 版 者 科学技术文献出版社
地 址 北京市复兴路15号 邮编 100038
编 务 部 (010) 58882938, 58882087 (传真)
发 行 部 (010) 58882868, 58882874 (传真)
邮 购 部 (010) 58882873
官 方 网 址 www.stdp.com.cn
发 行 者 科学技术文献出版社发行 全国各地新华书店经销
印 刷 者 虎彩印艺股份有限公司
版 次 2017年12月第1版 2017年12月第1次印刷
开 本 850×1168 1/32
字 数 32千
印 张 2.125
书 号 ISBN 978-7-5189-3359-4
定 价 22.00元



版权所有 违法必究

购买本社图书, 凡字迹不清、缺页、倒页、脱页者, 本社发行部负责调换

科技信息开放共享是国家创新发展的重要方向之一，也是国际科学社群高度关注的政策议题。对此，我国政府给予了高度重视。李克强总理在出席全球研究理事会 2014 年北京大会时指出，支持科技成果开放共享、保护支持产权及鼓励和培养青年科学家，是契合全球科学发展的现实需求和未来方向的。中国科学院及国家自然科学基金委员会也在 2014 年 5 月共同发布各自的公共资金资助科研成果的论文，实施开放获取的政策声明。

科技信息开放共享还意味着开放科研数据行动的实施。事实上，科研数据是数字化科研时代的重要战略资源，完善的数据共享政策可以为科研数据管理提供良好的指导，并且对于科研数据开放共享的实践具有积极的推动作用。早在 2003 年，德国、法国、意大利等国的科研机构就在德国柏林联合签署了由德国马普学会发起的《关于自然科学与人文科学资源的开放获取的柏林宣言》（Open Access to Knowledge in the Sciences and Humanities）。该宣言强调对于学术论文、教育资源、科研数据进行开放共享，呼吁各国科研机构向网络使



用者免费开放更多科学资源。经济合作与发展组织 (OECD) 于 2006 年颁布的《OECD 关于公共资助的科研数据获取的指导方针和原则》指出, 开放性意味着在同等条件下, 国际研究界以最低的成本——最好不超过传播的边际成本来获取科研数据。公共资助的科研数据的公开获取应该是容易获取的、及时的、对用户友好的, 并且最好是基于互联网的。英国皇家学会在 2012 年 5 月发布的《科学是开放事业》的报告中指出, 发布科学理论、实验和观察数据是其他人判断、同意、拒绝、理解该项工作的基础, 政府应制定科研数据开放政策, 并支持相关软件工具的开发和人员技能的培训等工作。开放科研数据是大势所趋, 国外相关机构已经制定了许多数据开放共享政策, 并且拥有相应的管理实践经验。

科学数据开放共享的积极措施由来已久, 其中, 基因组数据涉及国家安全、学术伦理、个人隐私、商业利益并且超越传统意义上的知识产权和开放获取的概念边界。美国国立卫生研究院 (NIH) 发布的一系列政策及颁布政策之前对各方意见征询的做法或可借鉴, 故整理编译有关信息。本书的初稿由郭进京、彭乃珠、宋忠惠进行编译, 顾立平进行校对和注释, 又在众多材料的基础上, 选择 NIH 具有业界影响力的政策文本进行编辑。感谢 NIH 公布并以知识共享许可的方式对文本进行传播。建议读者在 NIH 官网上查询这些政策内容并且进行适当引用。

Contents 目录

第 1 章 美国国立卫生研究院基因组数据共享政策序言	1
1.1 引言	1
1.2 范围和适用性	2
1.3 数据共享计划	4
1.4 非人类和模式生物基因数据	6
1.5 人类基因数据	7
1.6 研究人员获取和利用基因数据的责任	18
1.7 知识产权	21
1.8 总结	22
第 2 章 美国国立卫生研究院基因组数据共享政策	23
2.1 目标	23
2.2 范围和适用性	23
2.3 生效日期	25
2.4 提交基因组数据的研究人员的责任	25
2.4.1 基因组数据共享计划	25
2.4.2 非人类基因组数据	26
2.4.3 人类基因组数据	27
2.5 研究人员获取和利用基因组数据的职责	33



2.5.1 受控获取数据请求	33
2.5.2 受控获取数据的研究利用条件和协议	34
2.5.3 可自由获取数据的利用条件	36
2.6 知识产权	36
第 3 章 美国国立卫生研究院基因组数据共享政策	
补充信息	38
3.1 概述	38
3.2 有关 GDS 政策适用范围的指南	38
3.2.1 GDS 政策适用范围内的研究案例	38
3.2.2 GDS 政策适用范围外的研究案例	40
3.3 数据标准资源	40
3.4 数据提交和数据发布指南	41
3.4.1 0 级数据	43
3.4.2 1 级数据	44
3.4.3 2 级数据	44
3.4.4 3 级数据	45
3.4.5 4 级数据	46
第 4 章 美国国立卫生研究院基因组数据共享政策草案	
公众咨询评议会	47
4.1 公众评议会的流程概述	47
4.1.1 数据共享支撑着 NIH 的使命和发展优先级	47
4.1.2 数据共享政策扩展的动力	48



4.2 NIH 和数据共享：简要回顾历史文件和事迹	48
4.2.1 共享文化	48
4.2.2 需要更宽泛的政策	49
4.3 NIH 基因组数据共享政策草案概述	49
4.3.1 GDS 政策草案——主要构成	49
4.3.2 GDS 政策草案——范围和适用性	50
4.3.3 GDS 政策最终生效日期	51
4.3.4 GDS 政策草案——提交数据的研究人员的职责 ...	51
4.3.5 GDS 政策草案——非人类基因组数据	52
4.3.6 GDS 政策草案——知情同意预期	53
4.3.7 GDS 政策草案——数据提交机构的职责	54
4.3.8 GDS 政策草案——审议数据提交的 IRB 的职责 ...	54
4.3.9 GDS 政策草案——数据提交和数据发布预期	55
4.3.10 GDS 政策草案——获取 / 利用数据的 PI 的职责	56
4.3.11 GDS 政策草案——获取 / 利用数据的 PI 的职责（禁止事项）	57
4.3.12 知识产权	58

附录 A 美国国立卫生研究院基因组数据共享政策

管理框架图	59
-------------	----

美国国立卫生研究院基因组数据 共享政策序言

1.1 引言

美国国立卫生研究院 (National Institutes of Health, NIH) 宣布了最终的基因数据共享 (Genomic Data Sharing, GDS) 政策, 提出了确保基因研究数据广泛和可靠共享的期望。共享研究数据支持了 NIH 的使命, 而且对于促进研究成果转化为能够改善人类健康的知识、产品和程序也是必不可少的。除基因数据以外, NIH 的长期政策使大范围的研究数据能够及时公开并可用。

2013 年 9 月 20 日, NIH 在联邦登记册上发布了《NIH 基因数据共享政策草案征求公众评论》, 并于 2013 年 9 月 27 日发布了《NIH 捐赠和合同指南》, 实施为期 60 天的公众评论。

NIH 还利用网站、列表服务器及社会媒体传播征求意见



见的通知。在公众评论期间,2013年11月6日NIH举行了一次关于GDS政策草案的公开评议会,将近200人出席了会议,会议还包括一问一答部分。NIH一共收到了107个关于GDS政策草案的公众评论。

这些提交的评论来自个人、组织、学术机构附属单位、专业和科学社团、疾病和患者倡议团体、研究机构、行业和商业组织、部落组织、国家公共卫生机构和私人临床实践。

公众评论已经发布在NIH GDS网站上,这些评论都支持共享数据以促进研究的原则。但是,仍然存在一些问题和疑虑,并要求明确解释政策草案某些方面的内容。因而,我们对GDS政策相应章节进行组织评论,摘要如下。

1.2 范围和适用性

一些评论指出,该政策草案对于政策适用的研究类型还不明确。一些评论认为,一项研究中使用的技术(如基于阵列或高流量的基因技术)不应该成为决定政策适用性的重点。他们建议,从研究中获得的信息才应该是决定政策适用性的重点。

许多其他评论者表示,政策过于宽泛,这将导致大量提交的数据对其他研究者的实用性较低。其他一些评论者



认为,政策的范围还不够广泛。此外,一些评论者不确定政策是否适用于多种资金来源资助的研究。

NIH 已经修改了范围和适用性部分,以帮助明确该政策旨在应用的研究类型,而提及具体技术的内容已经被删除。研究项目类型示例的列表也在政策的范围内,可以参见 GDS 政策草案附录 A (现在被称为“NIH 基因数据共享政策的补充信息”),列表也进行了修订和扩充,不在范围内的研究示例也被增加进去了。

而且,最终的 GDS 政策现在明确规定,较小的研究(如测序少于 100 个研究参与者的基因)一般不适用于该政策。但是,规模较小的研究可能受制于其他 NIH 数据共享政策(如美国国家过敏和传染病研究所的数据共享和发布指南)或程序的要求。此外,对于该政策中使用的关键术语的定义(如汇总数据)也包括在内,其他条款也已被阐明。

关于范围的声明有意保持足够全面,以适应基因技术和产生基因数据的各种研究不断发展的本质。它也允许个别 NIH 研究所或中心(Institutes or Centers, IC)根据科学的状态、研究界的需要及 IC 的方案优先事项,选择以个案为基础应用的可能性,将该政策适用于产生小规模数据的项目。

该政策适用于 NIH 部分或全部资助的研究,前提是 NIH 资助支持基因数据的产生。研究人员对于政策是否适



用于他们当前或计划研究的疑问, 应该咨询相关的项目办公室或项目官员或 IC 的基因项目管理员 (Genomic Program Administrator, GPA)。GPA 的名称和联系方式可以通过 GDS 网站获得。

一些评论者表示担心研究人员和机构在验证和共享大量基因数据时带来的财政负担, 并且用于支持数据共享的资源将使资助远离研究。然而, NIH 坚持认为, 支持数据共享所需的资源并不是无价值的, 投资是必要的, 因为通过数据的二次使用可以使重要成果的发现成为可能。

此外, NIH 正在采取措施来评估和监测数据共享对科研行为成本的影响, 以编程的方式, 通过大数据分析到知识倡议, 并在组织上建立科学数据理事会, 这将对机构有关数据科学的问题提出建议。

1.3 数据共享计划

一些评论者指出, 关于在什么条件下 NIH 将对提交的基因数据授予例外的政策还不够清晰。有些人还建议, 当原始的同意书或者相关法律条文不允许广泛的共享时, NIH 应该允许人类基因数据的有限共享。

尽管 NIH 鼓励研究人员寻求广泛共享的一致性, 一些 IC 可能制订计划的优先事项, 期望研究提出资助, 以包括



同意广泛共享，也可以规定一些例外。最终，政策明确了当提交的基因数据不满足机构认证的标准时可能要求的免责条款。

一些评论者表示担忧：在一项研究完成之前，支持数据共享计划所需的资源将是难以估算的。其他人请求对要求用于支持数据共享计划的资源提供额外的指导。一些评论者认为，NIH 应该允许数据共享计划的某些要素可以随着其他及时的信息提交，如机构认证和相关的文档。对于多年度奖励，一些评论者建议，数据共享计划应该定期审查，以与当前道德标准一致。另外，一些人建议，数据共享计划应该公开。

根据 GDS 政策，研究人员将被期望在他们的资助申请预算部分列出他们需要准备提交给相应知识库资源的数据。如果有需要，NIH 将提供关于这些资源的额外指南。最终政策明确表示，在资助申请的资源共享计划部分，只有一个基本的基因数据共享计划需要提交资助申请，并需要在资助前提交一个更详细的计划。

机构认证应该在奖励之前提供任何其他及时的信息。基因数据共享计划指南可以在 NIH GDS 网站上获取：数据共享计划将通过年度进展报告或其他适当的科学项目审查来接受定期审查，并将进一步考虑数据共享计划应该公开的建议。



1.4 非人类和模式生物基因数据

GDS 政策草案提出了数据提交和数据发布的时间表 (例如, 当数据应该公开与其他研究人员共享时)。对于非人类数据, 该政策草案建议数据应该提交, 并在不晚于最初公布时间提供公开共享, 同时明确用于某些项目的数据提交和发布可能早于预期, 镜像数据共享预计已经在其他政策下到位。

一些评论者建议, 非人类数据的数据提交预期是不明确的。一位评论者建议, NIH 应该考虑一个比首次发布模型生物数据更快速的时间表, 同时其他评论者支持指定的数据发布时间表。而另有一些评论者则担心指定的时间表太短。

最终的 GDS 政策并没有改变提交和发布非人类和模式生物数据的时间表。时间表的推出基于促进广泛数据共享的需要, 同时还包含生成数据的研究人员, 他们往往必须做出显著的努力来准备用于共享的数据。该政策指出, 一个 NIH IC 可能选择缩短数据提交和发布特定项目的时间线, 并期望研究人员与 NIH 项目或项目官员就他们的项目时间表和里程碑的具体指南一起进行协调。

对于政策允许非人类和模式生物数据存储在任何广泛利用的数据知识库的灵活性具有普遍的支持。一位评论者



要求, 非 NIH 指定知识库的链接或引用应该包含到政策中。更多关于 NIH 指定知识库的信息, 包括这些知识库的例子, 可以在 GDS 网站查看, 关于非 NIH 指定数据知识库的额外信息将会纳入 NIH 工作人员和研究人员的宣传和培训资料中, 并在 GDS 网站公开。

NIH 已经明确了最终的政策, 并声明之前提交到广泛利用的知识库中的数据类型 (如基因表达数据到基因表达库或数组表示) 应该继续像以前一样, 同时之前没有提交的数据类型可能通过资助 IC 增加到这些或其他广泛利用的知识库中。

1.5 人类基因数据

NIH GDS 政策的补充信息基于该数据经历的处理水平的二次调查, 建立了提交和后续发布公开数据的时间表。一些评论者表达了对这些时间表的关注, 他们表示这些时间表太短, 可能限制了一个研究人员进行适当质量控制和在规定的时限内出版成果的能力。

许多评论者提出数据发布的时间表扩展到 12 个月或 18 个月, 或者到公布之日, 以先到之日为准的建议。其他人则担心, 目前的时间表太长, 他们应该反映长久以来快速的数据发布原则, 正如 Bermuda 和 Ft 阐明的那样。



Lauderdale 赞同一些评论者担心时滞期的取消 (如研究发布用于二次研究到提交的研究人员第一次出版研究成果之间的这段时间) 将对数据快速发布的目标产生不利的影响这一观点。一位评论者担心, 这些数据将在研究人员同参与人员讨论相应的发现之前公布。

NIH 已经修订了补充信息, 阐明不会启动对于 2 级和 3 级人类基因数据发布进行 6 个月的延期, 直到数据已经被清除。NIH 的数据提交政策已经启动, 一般发生在数据产生 3 个月之后。因为他们在研究项目中产生 2 级和 3 级人类基因数据将会有非常显著的差异, 提交的时间表根据具体的项目, 并将通过与资助的 NIH IC 研究人员协调, 由具体情况决定, 补充信息也已经对此做出了相应的说明。

在全基因组关联研究 (GWAS) 政策下, 出版时滞期被作为使数据更快速利用的一种方式。为了交换即时数据, 二级用户不允许出版或展现研究成果, 直到数据发布 12 个月。NIH 并没有为 GDS 政策采取这种方式, 因为在实践中, 出版时滞期对于二级用户来说很难追踪, 特别是对特定类型数据有多个不同时滞期的数据集, 这就增加了无意识违反时滞期的风险。

至于对人类基因数据将在研究人员通知参与人员有关重要发现之前公开的担忧, 这些数据将被认为是 4 级数据, 并且预计不会在出版前发布, NIH 坚信, GDS 政策将



为研究人员提供充分的时间同参与人员讨论相应的发现。

许多评论者呼吁该政策应该要求在提交人类基因数据时包含技术数据标准,如平台信息、控制词汇、规范算法、数据质量标准和元数据标准。NIH 明确了开发和利用基因数据标准的重要性,并认识到有许多正在开发和推广这类标准的倡议。

NIH 已经通过增加数据标准资源部分修订了补充信息。它为数据提交到具体的 NIH 指定数据知识库中提供了参考,包括数据标准。有关数据标准的其他资源将被合并到补充信息中,因为这能使得它们被开发并变得适合广泛地应用。

一些评论者要求定义 NIH 指定的数据知识库及判定哪些非 NIH 知识库是可以作为这种知识库被接受的指南。评论者也表示了对利用受信任合作伙伴详细信息的兴趣,通过合同机制建立第三方合作伙伴关系,从而满足为那些对基因数据分析有用数据的存储和 / 或使用相关工具提供基础设施服务的需求。

关于 NIH 指定知识库的定义,目前已经包括在最终的政策中。此外,关于接受人类基因数据的非 NIH 指定知识库的详细信息将在 NIH GDS 网站上公布,并将其纳入 NIH 工作人员和 NIH 资助研究人员的宣传和培训材料中。

关于信任的合作伙伴的更多信息,包括信任的合作伙伴要