

● 宋筱璇¹, 王延飞^{1,2}, 钟灿涛³

(1. 北京大学 信息管理系, 北京 100871; 2. 北京大学“一带一路”研究中心, 北京 100871;
3. 北京大学先进技术研究院, 北京 100871)

国内外科研数据安全政策比较研究*

摘要: 在数据共享热潮涌动的当下, 科研数据安全问题日渐突出。针对各科研机构对共享中数据安全问题的政策原则及处理方式展开情报调研, 发现国外在科研数据共享中的安全意识较高, 管理政策较完善; 国内在数据识别、评估监管及数据处理等方面仍存在较大的政策空白。建议我国在科学交流管理中应尽快制定和完善数据安全政策与相关规定。

关键词: 数据共享; 科研数据; 信息安全; 数据政策

Abstract: With the blossoming of data sharing, security issues about research data are becoming increasingly prominent. In this paper, we survey policies or regulations and practices related to data sharing and data security in scientific institutions. It is impressed that relatively high security awareness and complete management policies exist in research data sharing abroad, whereas it is insufficient for Chinese scientific institutions in the areas such as data identifying, assessment management and data alignment. It is extremely significant for China to develop and to improve data security policies and regulations in the management of scientific communication as soon as possible.

Keywords: data sharing; research data; information security; data policy

科研活动需要数据的支撑。研究过程中, 为获取数据往往需要消耗大量的人力、物力, 同时也造成了数据的浪费。在这种情况下, 科研数据的共享开始受到广泛的关注。然而, 在研究人员欣喜于科研数据共享的成果时, 数据安全问题日益暴露出来。随着大数据的到来, 数据安全问题被愈加放大。如何既确保安全, 又能高效地共享数据? 对科研数据中的敏感信息如何处理? 如何保护科研活动中的隐私? 这些都是亟待被思考的问题。

1 相关研究

科研数据, 也称为科学数据, 是科研活动中重要的产出之一^[1]。在开放共享的大潮中, 科研数据在科学合作、传播和交流中具有重大战略意义。科研数据安全既包括科研数据本身的安全, 也包括数据相关利益主体的安全, 特别是涉及主体隐私权的数据, 在共享中受到更多的关注^[2]。

利用 CNKI 和 Web of Science 文献数据库分别对国内外科研数据安全研究相关的文章进行检索发现, 尽管国内

外都存在大量数据安全相关的研究, 但国内对于直接研究科研数据安全的文献还相对较少。在 CNKI 中以“科研数据安全”为检索词进行主题检索, 仅获得 4 篇相关文献, 且主要研究数据平台或系统的安全保障问题。而对于科研数据安全问题的探讨往往嵌入在科研数据共享、开放获取、数据管理等政策或技术研究中。

在科研数据管理或科研数据开放共享研究中, 国内的相关研究基本处于介绍国外政策实例阶段, 科研数据安全问题只是作为其组成部分, 通过经验案例进行介绍, 缺乏与我国本土实际情况的联系。张瑶等^[3]在调研与分析国外科研资助机构数据政策后发现, 科研资助机构的数据政策中往往涉及数据的保密及安全规定。资助机构要求研究者遵循相关数据保护法规, 特别要对机密和隐私数据设定数据共享限制, 对人类受试者数据应采取匿名化、重新鉴定、隐藏标识符等方式处理。丁培^[4]在对国外大学科研数据管理政策研究时发现, 美国、英国以及澳大利亚的许多大学在数据管理政策中都涉及数据安全、隐私和保护的内容, 具体包括: 数据安全保护及保密的原则性说明; 关于数据限制性访问的说明; 隐私以及保密协议的说明。张闪闪等^[5]在对国外信息服务机构的数据管理政策进行调研时也发现, 斯坦福大学图书馆在科研数据的处理上, 注意隐私保护。当涉及人类受试者数据时, 要接受监督并使用

* 本文为国家社会科学基金一般项目“创新驱动战略的情报保障研究”的阶段成果之一, 项目编号: 16BTQ058。

专门的数据搜集工具。

相较于国内,国外对于科研数据安全政策及技术的研究则更为全面,特别是在科研数据共享中涉及的隐私及敏感数据等问题受到广泛关注。在政策上,美、英等国的科研资助机构(美国国家科学基金会(NSF),英国研究理事会(RCUK)等)和高校(哈佛大学、牛津大学等),纷纷制定了各自的科学数据开放政策指南或规定,其中包括数据的分类规则、数据共享的限制、数据处理等与科研数据安全相关的重要内容。在技术上,针对不同领域数据的特点,为数据存储、访问等过程的安全问题提供保障。Sydes等^[6]基于临床医学试验数据共享中的信息披露风险,开发了一套新的访问控制的方法,从而保障数据共享中的安全。Dong等^[7]对大数据共享平台中涉及用户个人信息的敏感数据,提出了新的基于异构密文转换代理算法和基于虚拟机监视器的用户进程保护方法,为安全共享这些敏感数据提供支持和保障。在涉及隐私及敏感数据的讨论中,共享数据的识别或评价标准问题一直是相关学者关注的热点之一,但始终没有一套相对通用的评估方法。如Sariyar等^[8]意识到帮助数据提供者来识别科学数据共享中可能存在的法律、伦理或社会敏感问题是一个经常被忽略但却十分重要的环节。同时描述了一套复杂的监管环境,并提供相应的在线工具帮助数据供应商识别敏感数据及数据中涉及的法律伦理问题。Malin等^[9]改进了安全港的HIPAA标准隐私规则和传统的基于统计标准的隐私规则模型中的不足,提出了一个更直观的De-identification方法对医疗研究中的病例数据进行评估,通过评估结果确认数据是否达到共享标准。

此外,随着近年来科研数据共享中的数据安全受到越来越广泛的重视,对于数据“边开放边保护”的呼声越来越高,国外对于数据共享与数据安全之间关系的博弈也出现了新的现象。Dove^[10]认为,对数据采用复杂的再编码方式可能仅仅只是给机构审查委员会和研究伦理委员会创造的一种虚假的安全感,在一定程度上会阻碍科研数据的再研究和创造。Stoddart等^[11]发现欧盟的数据保护是以特定地区法律或政策作为标准,对数据的透明度、质量、比例、安全、访问和修编等限制方面进行评估。理论上,每个国家或地区的数据保护规定与欧盟的数据保护原则应该是一致的或相似的。然而,近年来在数据保护决策方面的不一致愈发明显,仅有5个国家或地区与欧盟相一致。这样的不一致产生的直接负面作用就是阻碍了欧盟成员之间的数据共享。因此,在科研数据共享中,意识到数据安全问题,从政策和技术上为数据共享中的安全提供保障支持固然重要,但如何更好地实现“管为用,用中管”,在科研数据共享和安全中维持良好的平衡也是未来

研究重要的方向。

近年来,我国各部委、省市及高校等科研单位纷纷建立起科学数据共享平台。数据共享中的安全管理政策制定存在较大的优化空间。本文拟从科研数据共享与安全的基本原则、启动时机、数据识别、评估监管和结果处理5个方面(如图1所示),对比国内外各类科研机构或科研数据共享平台的数据安全政策,并分析其中的共性与不足,以期为我国科研共享活动中的数据安全政策的规划和建设提供相应的借鉴。

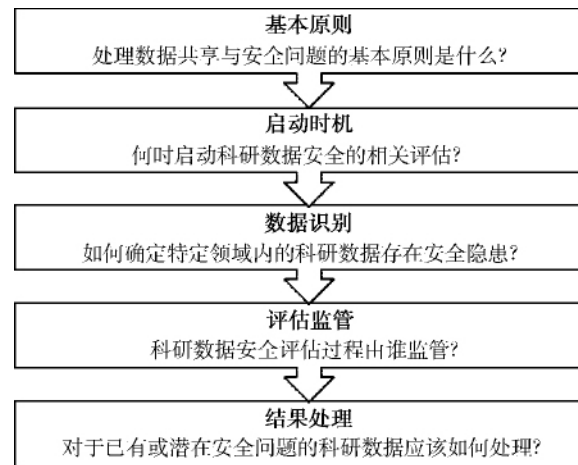


图1 科研数据共享与安全管理政策比较框架

2 对国内外科研数据管理政策的解析

从国家基金组织、科研资助机构和高校等领域入手对国内外科研数据共享与安全的管理政策进行分析,可以发现国外呈现“鼓励共享,安全协同”的状态,国内则呈现“鼓励共享,安全不足”的状态。

2.1 “鼓励共享,安全协同”的国外科研数据管理政策

1) 国家基金组织、科研资助机构。根据DCC^[12]网站上发布的英国研究资助者数据管理和共享计划文件汇总,选取在数据管理和共享政策方面较为完善的6所科研资助机构,分别是:英国研究理事会(RCUK)^[13]、英国癌症研究中心(CRUK)^[14]、英国生物技术与生物科学研究理事会(BBSRC)^[15]、英国经济和社会研究理事会(ESRC)^[16]、英国医学研究理事会(MRC)^[17]、英国环境研究委员会(NERC)^[18];同时纳入美国国家科学基金委(NSF)^[19]和美国国立卫生研究院(NIH)^[20]共计8所科研资助机构作为分析对象。

英国癌症研究中心(CRUK)^[14]在其“数据共享和保护策略”中提出该条例不规定研究者应该何时以及如何保护和共享数据,而是使研究者在计划和进行研究时明确应该注意的问题。调研中也发现,除了美国国立卫生研究

院 (NIH)^[20] 就人类受试者的隐私和数据保护问题进行了相对具体的规定, 其余基金组织及科研资助机构主要着眼于从整体政策的角度探讨科研数据共享与安全中可能存在

的问题, 而不涉及具体数据的识别、处理和监管细则。8 所国外组织机构中涉及数据共享中数据安全问题的代表性文件及政策细则如表 1 所示。

表 1 国外基金组织、科研资助机构数据共享与安全政策条例一览

科研资助机构名称	相关文件	出台时间	政策细则 (涉及数据共享中的数据安全问题)
英国研究理事会 (RCUK)	数据政策通用原则	2011	1. 最大化数据共享机会; 防止参与者的权益因不适当的公开数据而受损; 2. 关注法律、伦理和商业中对研究数据共享的限制; 3. 数据共享限制问题应被关注在研究开始前
英国癌症研究中心 (CRUK)	数据共享和保护	2009	1. 及时、负责任地共享数据, 包括那些与伦理有关的数据; 2. 对于涉及个人数据的医学研究, 应有适当的监管权限; 3. 申请时应提交数据管理和共享计划
英国生物技术与生物科学研究理事会 (BBSRC)	数据保护政策	2012	1. 关注个人数据的保护, 依据隐私规则使用数据的过程必须是透明的, 避免不恰当地共享和使用数据给个人带来不利的影响; 2. 在机构中设立数据保护专员、信息和记录管理者对数据中涉及的隐私影响进行评估及对数据共享进行指导; 3. 申请时应提交数据管理和共享计划
英国经济和社会研究理事会 (ESRC)	研究数据政策	2014	1. 追求最大化的数据共享, 但承认法律、伦理或商业约束数据共享的存在, 这些限制应该被详细考虑在研究开始之前, 并贯穿研究数据的整个生命周期中; 2. 数据依据其细节、敏感性和机密性水平进行分类。在此分类基础上进行数据处理和数据安全保障的控制; 3. 申请时应提交数据管理和共享计划
英国医学研究理事会 (MRC)	人口患者研究数据共享政策和指导	2012	1. 提出对于研究中的敏感数据应建立一套安全管理及使用标准
英国环境研究委员会 (NERC)	数据政策	2016	1. 共享数据必须符合信息自由和环境信息法规, 当数据的共享会给环境带来破坏时应该被限制
美国国家科学基金委 (NSF)	研究成果的传播和分享	2011	1. 鼓励和促进数据共享, 特权或机密数据在发布形式上应以保护个人隐私和主题为原则; 2. 访问和共享数据时应进行适当的隐私保护, 注意数据的机密性和安全性; 3. 申请时应提交数据管理计划补充文档
美国国立卫生研究院 (NIH)	数据共享政策和实施指导	2003	1. 尽可能广泛、免费地对所有数据进行共享, 同时保护参与者的隐私以及专有数据的机密; 2. 关注人类受试者的隐私和数据保护问题, 人类受试者的数据应遵守 HIPPA 隐私规则, 受机构审查委员会 (IRB) 的监管; 3. 在数据共享前, 应进行数据匿名处理, 重新修订所有标识符, 以减少信息披露的风险; 4. 申请时应提交数据共享计划

从表 1 的结果可知, 这些机构普遍具有与数据共享与安全相关的政策或规定, 但是这些规定大都以“数据”作为其命名主要成分, 没有将“数据安全”问题作为独立个体进行规范, 数据安全问题仅仅是其政策规定的部分内容。

这些机构在数据共享与安全问题的基本原则上都鼓励共享, 如 RCUK 的“最大化共享”, CRUK 的“及时、负责任的共享”, NIH 的“广泛、免费的共享”等, 也都承认存在数据安全或数据隐私的问题, 这些问题可能包括数据的机密性、敏感性, 受试者的隐私等各类法律伦理限制, 而当研究涉及人类时, 数据的安全问题就显得尤为重要。NIH^[20] 规定涉及人类受试者的数据应遵守 HIPPA 隐私规则, 受机构审查委员会 (IRB) 的监管; 同时规定在数据共享前, 应进行数据匿名处理。具体表现为重新修订

所有标识符, 以减少信息披露的风险。对于姓名、地址、电话号码和社会安全号码一类的直接标识符应直接删除; 对于间接标识符和其他信息应防止“演绎披露”的发生。(所谓“演绎披露”即通过一些变量的联系推测出一些关键的变量特征。如将人口信息、地理信息和其他信息联系起来可能反映出受试者的某些隐私信息。) 在数据共享中可采用两种方式来确保数据安全: 仅保留数据的一部分内容; 或是从统计学的角度对数据进行二次编码。

在数据安全相关问题的关注时间上, RCUK、CRUK、NSF、NIH 等科研机构都主张任何有关数据的问题都应该在研究开始前就先被考虑到。ESRC 承认法律、伦理或商业约束会影响数据共享, 提出这些限制应该被详细考虑在研究开始之前, 并贯穿研究数据的整个生命周期中。此外, 调研发现 8 个科研资助机构都规定研究者在申请项目

时应该同时提交一份类似于“数据管理和共享计划”的文件。该文件作为申请的评估对象之一描述的是申请者对申请项目研究成果共享中应该遵守的政策规定。例如 NSF^[19] 的数据管理和共享计划要求包括以下 5 方面的内容：①研究中的数据、样本、物理集合、软件、课程材料和项目过程中产生的其他材料的类型；②数据的标准，元数据的格式和内容；③适当的隐私保护，数据机密性，安全性，知识产权或其他权利或要求；④对重用、再分配及项目衍生品生产的政策或规定；⑤对数据、样本和其他研究产品的归档及访问保护计划。

从数据的识别和评估来看，尚未有政策直接对如何识别敏感数据进行规定。但 ESRC 依据其细节、敏感性和机

密性水平对数据进行分类，在此分类基础上进行数据处理和数据安全保障的控制，因而从实质说来，数据分类的过程就是数据识别或评估的过程。

2) 高校。高校作为研究实体，在依据各科研资助机构政策规则的基础上，对实际数据的识别和敏感数据的处理方面进行了更详尽的规定，尽可能地将存在共享限制或潜在威胁的数据转化为可以公开的数据，在更大程度和范围上实现科研数据的共享。本文选取了哈佛大学^[21]、麻省理工学院^[22]、加州大学伯克利分校^[23]、斯坦福大学^[24]、牛津大学^[25] 5 所国外优秀高校，考察其已有的科研数据共享与安全的管理政策，特别关注其中关于数据的识别、监管和处理细节，具体如表 2 所示。

表 2 国外高校科研数据分类及处理政策一览

学校名称	数据分类	限制共享数据类型	数据评估参考	数据处理
哈佛大学	依据数据披露后造成的危害程度分为 5 级数据	2~5 级数据	机构审查委员会 (IRB) 遵守 HIPPA 隐私规则 胚胎干细胞研究监管委员会 (ESCRO) 基因组数据政策 (GDS) 微生物安全委员会 (COMS) 等	不同级别的数据采用不同的数据调整办法
麻省理工学院	——	人类受试者数据、医疗健康数据	遵循 ICPSR 指南 遵守 HIPPA 隐私规则 人类被试委员会 (COUHES)	数据匿名化
加州大学伯克利分校	基于数据需要的保护水平，分为 0~3 共 4 个级别	1~3 级数据	保护人类受试者委员会 (CPHS)	不同保护水平的数据采取不同的存储和共享方式
斯坦福大学	——	人类受试者数据、医疗健康数据、敏感数据	机构审查委员会 (IRB) 遵守 HIPPA 隐私规则	数据调整
牛津大学	——	机密数据、人类受试者数据、敏感数据	研究伦理委员会 (REC)	数据匿名化

注：“——”表示相关政策或规定中不包含此项内容。

从表 2 可以看出，在数据评估与识别上，部分高校对其科研数据类型进行了划分。其中哈佛大学为解决各领域机密和敏感数据的保护问题，在其数据安全政策 (HRD-SP)^[21] 中，将不同领域的数据依据其安全控制需求的程度分为 5 个级别，分别是：①非机密研究信息；②应该设为机密的信息；③敏感或者机密的信息；④非常敏感的信息；⑤极度敏感的信息。绝大多数涉及人类受试者的敏感数据被划分为第 3 级数据，与国家安全相关的数据通常属于第 4 级数据。HRDSP 还规定了每个级别数据的最低保护要求。研究人员在项目初始之时，首先需要为即将产生的数据确立一个安全级别，以确定研究中对产生数据的安全控制水平。研究者可自行确定研究数据 1~3 级安全级别，但若研究涉及 4、5 级数据，则需要提交哈佛大学信息技术委员会 (HUIT) 审批。加州大学伯克利分校^[23] 基于数据需要保护的级别，制定了一套数据分类标准，以此评估数据的敏感性，具体分为 0~3 共 4 个级别。0 级：公开的信息，例如网页、课程信息等；1 级：在一定条件下

可发布的信息，例如学生学号；2 级：有一定保密要求的数据元素，例如社会保险号；3 级：在多个敏感系统之间存在共享风险的数据，例如备份数据系统。此外，未对数据进行分类的高校，共享中被限制较多的数据主要集中在人类受试者数据、医疗健康数据、敏感数据和机密数据几部分。

在数据监管责任者方面，国外的高校主要集中在对人类受试者数据和医疗健康数据设有相应的评估机构和评估规范来确定该数据是否满足共享的标准或是否触犯相应的限制。哈佛大学和斯坦福大学都设有机构审查委员会，麻省理工学院设有被试委员会，加州大学伯克利分校设有保护人类受试者委员会，牛津大学设有研究伦理委员会对人类受试者数据进行评估。哈佛大学、麻省理工学院和斯坦福大学等还规定人类受试者和医疗健康数据必须遵守 HIPPA 隐私规则。这些机构或规则都以保护科研活动中的人类受试者为主要目的。以哈佛大学的机构审查委员会为例，下设人类受试者使用委员会 (CUHS) 和人类研究管

理办公室 (OHRA) 分管不同的项目。当研究涉及与人相关的新药物或疗法的调查、观察研究、人体组织研究或其他关于人类的数据研究时, 都必须在项目开始之前向 IRB 的 ESTR 平台提交申请, 申请包括其项目的完成形式和示例文档, 由 IRB 对应办公室人员进行 4 周左右的时间审核, 返回审核结果之后才能启动研究项目。有时, 一个研究项目可能需要多个委员会共同审核及批准, 例如当进行干细胞研究时, 哈佛大学就规定需要经 IRB 和胚胎干细胞研究监察委员会的共同审核。除了对人类受试者和医疗健康数据进行监管外, 对于其他领域的的数据, 各高校也设立了对应的责任机构, 例如哈佛大学在处理生物数据安全的相关问题时, 通过微生物安全委员会的监管来进行。

在数据处理方面, 目前主要采用数据匿名或数据调整的方式对敏感数据进行处理。数据匿名和数据调整方法是对科研数据中与研究对象身份相关的直接标识符进行删除处理, 部分间接标识符进行重新编码后有选择地共享, 但处理的力度可能依据数据分类的不同级别而有所不同。对于可将变量之间相互关联, 从而能透露身份信息的关系数据应受到特别的关注。但数据处理也不宜过度, 牛津大学指出, 应实现合理化水平的数据匿名, 不可不切实际或过于严厉地处理数据, 当研究数据是一些采访转录时, 粗暴地删除或聚合标识符, 将使数据扭曲无法使用, 可使用替代或模糊描述的方式加以转化。

3) 期刊、数据库。除了上述机构和高校外, 还有一些与科学研究密切相关的机构组织也关注科研数据的共享与安全问题, 期刊和数据库就是其中的代表。

《自然》^[26] 提出对出版论文及其数据访问中可能造成的社会影响所产生的担忧, 这其中包括伦理问题、生物安全问题等。特别指出由于生化武器威胁的存在, 对于生物论文及其数据的潜在安全威胁应由专家学者进行风险评估, 并应有一个合适的政策来处理此类问题。尽管如此, 在遵循共享与安全问题处理的基本原则基础上, 编辑部仍是鼓励共享的。2012 年 2 月《自然》期刊编辑部提出“尽管尚存争议, 但发布敏感数据的好处远大于不出版它们所面临的风险”^[27], 因而决定将一篇流感的论文全部出版。虽然面临其中流感病毒的数据可能被恐怖组织或是个人疯狂行为滥用的风险, 但是该数据的发布对流感病毒的监测和人体健康具有重大意义。

美国校际社会科学数据共享联盟 (Inter-university Consortium for Political and Social Research, ICPSR)^[28] 是现在世界上最大的社会科学数据中心, 在收集社会科学数据的过程中, 也特别关注对受试者身份的保护。ICPSR 规定在数据共享之前应将信息重新编码, 包括将日期转换为时间间

隔, 将出生日期转换为年龄组, 将详细的地理信息转换为更广泛的地理位置, 将收入转换为收入类别等。将被试的标识分为直接标识符和间接标识符两类, 对于显性地指向特定的个人或单位的直接标识符在数据共享前必须进行移除或掩盖。

2.2 “鼓励共享, 安全不足”的国内科研数据管理政策

与国外在科研数据共享与安全问题上的“鼓励共享, 安全协同”管理政策有所不同, 国内在科研数据的管理中整体上呈现“鼓励共享, 安全不足”的状态。

我国近年来建立了多个与数据共享相关的平台。从国家角度, 为加强科技创新基础能力建设, 推动我国科技资源的整合共享与高效利用, 按照《关于开展国家科技基础条件平台认定和绩效考核工作的通知》(国科发计 [2011] 318 号) 要求, 科技部与各大高校、研究机构合作, 完成了首批 23 家国家科技基础条件平台认定评审工作, 并于 2011 年 11 月 9 日向全社会公布。其中包括 6 个科学数据共享平台, 分别是: 林业科学数据平台^[29]、地球系统科学数据共享平台^[30]、人口与健康科学数据共享平台^[31]、农业科学数据共享中心^[32]、地震科学数据共享中心^[33]、气象科学数据共享中心^[34]。各省市地方响应号召, 建立起各自的科学数据共享平台, 目前已完成的包括陕西省科学数据共享平台^[35]、山西省科学数据共享平台^[36]。此外, 各高校和科研院所也逐步建立起自己的科研数据共享平台或中心, 例如北京大学^[37] 设立的开放研究数据平台, 中国科学院的基础科学数据共享网^[38], 等等。

本文选取科技部的建构较为完善的 6 个科学数据共享平台, 来了解国内科研数据共享与安全管理政策的大致发展情况, 具体如表 3 所示。

从表 3 中可以看出, 林业科学数据平台、人口与健康科学数据共享平台、农业科学数据共享中心和地震科学数据共享中心都已经意识到共享数据存在安全问题, 特别是涉及国家安全的涉密数据的安全问题。除此之外, 还有个人隐私等法律限制方面的问题, 并将具体规定写入各自的数据共享管理办法或数据共享元数据标准中。

数据分类可以在一定程度上为共享数据的限制评估做准备。在数据分类上, 地震科学数据共享中心根据地震科学数据发布和共享的范围, 将数据划分为以下 4 级: 1 级数据, 凡可向社会公众公开发布的数据; 2 级数据, 能够向国内、国外用户提供的数据; 3 级数据, 可以向国内用户提供的数据; 4 级数据, 只允许向特定范围的用户提供的数据。并规定用户使用 1 级数据, 可以在地震科学数据共享服务机构的网站上浏览、查询和下载; 用户使用 2 级和 3 级数据, 应在地震科学数据共享服务机构的网站上完成相应的注册程序后获得, 必要时也可通过签订合同的方式

表3 科技部科学数据共享平台数据安全政策一览

数据共享平台	相关文件	数据分类	限制共享数据类型
林业科学数据平台	林业科学数据元数据标准	绝密、机密、秘密、限制、内部、无限制	涉及国家安全、个人隐私、知识产权以及其他特殊要求
地球系统科学数据共享平台	——	——	——
人口与健康科学数据共享平台	人口健康科学数据共享元数据标准	绝密、机密、秘密、限制、内部、无限制	法律限制(保护隐私权或知识产权)和安全限制分级(从国家、单位或个人安全考虑)
农业科学数据共享中心	农业科学数据共享管理办法	——	涉密数据
地震科学数据共享中心	地震科学数据共享管理办法	4级数据	涉密数据、2~4级数据
气象科学数据共享中心	——	——	——

注：“——”表示该平台中不包含此项内容。

式获得；用户使用4级数据，应向地震科学数据共享服务机构提出申请，并经审核后方可获取所需数据。林业科学数据平台和人口与健康科学数据共享平台则根据《GB/T 7156—1987 文献保密等级代码》，将数据划分为6个保密级别，分别为公开数据、国家内部数据、部门内部数据、秘密数据、机密数据、绝密数据。

总体看来，国内的部分数据共享平台虽然能够意识到数据安全问题的存在，也有一定的数据管理政策或规定，但其对于数据类型的划分或者是限制共享数据的识别的规定还过于宽泛，缺乏评估或识别的标准，也缺乏在数据评估执行中的责任或监管机构。同时，对于存在共享限制但在一定条件下能够共享的数据，目前也没有一套公开的数据处理管理办法。

3 总结和建议

通过对比国内外科研机构数据共享与安全管理的相关政策规定可以发现，目前，国外主要采用“鼓励共享，安全协同”的科研数据管理政策，对科研数据共享中的数据安全意识较高，管理架构相对完善。国内也开始意识到数据安全相关的问题，但仍处于探索发展阶段，呈现“鼓励共享，安全不足”的状态。基于我国共享中的数据安全政策存在的不足，本文提出以下几点建议：

第一，建立一套相对完整的数据安全问题评估标准。虽然部分数据共享平台尝试进行数据分类，但此种分类主要围绕数据的保密分级进行，与数据的隐私、健康信息识别、数据的安全敏感度或潜在威胁识别等要求尚有不小的距离。在评估规范中落实关注安全隐患、对潜在威胁进行评估，才有可能识别出共享交流中存在安全问题的数据。

第二，组建跨领域的数据共享监管责任团队。目前，国内的数据共享平台中，数据使用及共享管理仅对本平台负责，缺乏跨域、跨平台的安全评估或监管介入，留出对科研数据共享的监管漏洞。

第三，权衡处理好数据安全与共享的关系问题。数据

调整和数据匿名的方式，我们可以借鉴，但应慎重把握数据共享和数据安全中的度。过度调整虽然在一定程度上保护了数据，但也在更大程度上阻碍了数据的共享。□

参考文献

- [1] 姜鑫. 科学数据开放政策研究现状分析及未来研究动向评判[J]. 现代情报, 2016, 36(2): 167-170.
- [2] 马海群, 蒲攀. 国内外开放数据政策研究现状分析及我国研究动向研判[J]. 中国图书馆学报, 2015(5): 76-86.
- [3] 张瑶, 顾立平, 杨云秀, 等. 国外科研资助机构数据政策的调研与分析——以英美研究理事会为例[J]. 图书情报工作, 2015, 59(6): 53-60.
- [4] 丁培. 国外大学科研数据管理政策研究[J]. 图书馆论坛, 2014(5): 99-106.
- [5] 张闪闪, 顾立平, 盖晓良. 国外信息服务机构的数据管理政策调研与分析[J]. 图书情报知识, 2015, 167(5): 99-109.
- [6] SYDES M R, JOHNSON A L, MEREDITH S K, et al. Sharing data from clinical trials: the rationale for a controlled access approach[J]. Trials, 2015, 16(1): 1-6.
- [7] DONG X, LI R, HE H, et al. Secure sensitive data sharing on a big data platform[J]. Tsinghua Science & Technology, 2015, 20(1): 72-80.
- [8] SHANE S, SHELAGH M C, SHELLEY R B, et al. Sharing and reuse of sensitive data and samples: supporting researchers in identifying ethical and legal requirements[J]. Biopreservation & Biobanking, 2015, 13(4): 263-270.
- [9] MALIN B, BENITEZ K, MASYS D. Never too old for anonymity: a statistical standard for demographic data sharing via the HIPAA Privacy Rule[J]. Journal of the American Medical Informatics Association, 2011, 18(1): 3-10.
- [10] DOVE E S. Biobanks, data sharing, and the drive for a global privacy governance framework[J]. Journal of Law Medicine & Ethics, 2015, 43(4): 675-689.
- [11] STODDART J, CHAN B, JOLY Y. The European Union's adequacy approach to privacy and international data sharing in

- health research [J]. The Journal of Law , Medicine & Ethics , 2016 , 44.
- [12] DCC. Summary of UK research funders' expectations for the content of data management and sharing plans [EB/OL]. [2016-06-25]. http://www.dcc.ac.uk/sites/default/files/documents/resource/policy/FundersDataPlanReqs_v4%204.pdf.
- [13] RCUK. RCUK common principles on data policy [EB/OL]. [2016-06-25]. <http://www.rcuk.ac.uk/research/datapolicy/>.
- [14] CRUK. Submission of a data sharing and preservation strategy [EB/OL]. [2016-06-25]. http://www.cancerresearchuk.org/sites/default/files/data_sharing_policy.pdf.
- [15] BBSRC. BBSRC data sharing policy [EB/OL]. [2016-06-25]. <http://www.bbsrc.ac.uk/documents/data-sharing-policy-pdf/>.
- [16] ESRC. ESRC research data policy [EB/OL]. [2016-06-25]. <http://www.esrc.ac.uk/files/about-us/policies-and-standards/esrc-research-data-policy/>.
- [17] MRC. MRC policy and guidance on sharing of research data from population and patient studies [EB/OL]. [2016-06-25]. <http://www.mrc.ac.uk/research/policies-and-resources-for-mrc-researchers/data-sharing/data-sharing-population-and-patient-studies/>.
- [18] NERC. NERC data policy-2016 [EB/OL]. [2016-06-25]. <http://www.nerc.ac.uk/research/sites/data/policy/datapolicy-guidance/>.
- [19] NSF. Dissemination and sharing of research results [EB/OL]. [2016-06-25]. <http://www.nsf.gov/bfa/dias/policy/dmp.jsp>.
- [20] NIH. NIH data sharing policy and implementation guidance [EB/OL]. [2016-06-25]. http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm.
- [21] Harvard research data security policy (HRDSP) [EB/OL]. [2016-06-25]. <http://vpr.harvard.edu/pages/harvard-research-data-security-policy>.
- [22] Data management [EB/OL]. [2016-06-25]. <http://libraries.mit.edu/data-management/>.
- [23] Data classification standard [EB/OL]. [2016-06-25]. <https://security.berkeley.edu/data-classification-standard#Protection-Level-2>.
- [24] Sharing sensitive data [EB/OL]. [2016-06-25]. <http://library.stanford.edu/research/data-management-services/share-and-preserve-research-data/sharing-sensitive-data>.
- [25] To share or not to share? [EB/OL]. [2016-06-25]. <http://researchdata.ox.ac.uk/home/sharing-your-data/to-share-or-not-to-share/>.
- [26] Biosecurity policy [EB/OL]. [2016-06-25]. <http://www.nature.com/authors/policies/biosecurity.html>.
- [27] Flu papers warrant full publication [EB/OL]. [2016-06-25]. <http://www.nature.com/nature/journal/v482/n7386/full/482439a.html>.
- [28] Confidentiality [EB/OL]. [2016-06-25]. <http://www.icpsr.umich.edu/icpsrweb/content/datamanagement/confidentiality/>.
- [29] 林业科学数据平台 [EB/OL]. [2016-06-25]. <http://www.cfsdc.org/>.
- [30] 地球系统科学数据共享平台 [EB/OL]. [2016-06-25]. <http://www.geodata.cn/>.
- [31] 人口与健康科学数据共享平台 [EB/OL]. [2016-06-25]. <http://www.ncmi.cn/1>.
- [32] 农业科学数据共享中心 [EB/OL]. [2016-06-25]. <http://www.agridata.cn/>.
- [33] 地震科学数据共享中心 [EB/OL]. [2016-06-25]. <http://data.earthquake.cn/policy/gxbf.htm>.
- [34] 气象科学数据共享中心 [EB/OL]. [2016-06-25]. <http://data.cma.cn/>.
- [35] 陕西省科学数据共享平台 [EB/OL]. [2016-06-25]. <http://www.snsd.gov.cn/Portal/?isCookieChecked=true>.
- [36] 山西省科学数据共享平台 [EB/OL]. [2016-06-25]. <http://kxsj.sxinfo.net/>.
- [37] 北京大学开放研究数据平台 [EB/OL]. [2016-06-25]. <http://opendata.pku.edu.cn/about.xhtml>.
- [38] 中科院基础科学数据共享网 [EB/OL]. [2016-06-25]. <http://www.nsde.cn/>.
- 作者简介: 宋筱璇, 女, 1992年生, 硕士生。研究方向: 情报研究。
王延飞, 男, 1965年生, 博士, 教授, 博士生导师。研究方向: 情报研究, 决策信息保障。通讯作者。
钟灿涛, 男, 1969年生, 博士, 副研究员。研究方向: 科研管理。
- 收稿日期: 2016-06-05