英国科学数据管理概述

王 静[1]，马慧勤[2]
（1. 中国科学技术部，北京 100862；
2. 中国科学技术交流中心，北京 100045）

摘 要：科学数据开放共享近年成为开放获取的重点。英国率先提出 e-science 概念，在科学数据管理和共享方面有很多实践经验。本文概述了英国科学数据管理情况，包括管理与共享政策、科学数据中心建设、国家财政支持的科技项目产生的科学数据管理与共享机制、科学数据保密与安全管理等情况，并对我国科学数据管理提出发展建议。

关键词：英国；科学数据；数据管理；数据共享
中图分类号：G327.561 **文献标识码**：A DOI：10.3772/j.issn.1009-8623.2018.06.006

对于科研产出的开放获取（Open Access），以往的关注点多在学术论文、报告等正规出版物的公开获取，近年来科学数据的开放共享成为关注的重点。英国在 2000 年率先提出了 e-science 概念，后被各国广泛采用并进一步发展，主要指大气与地球科学、环境科学、天文学、高能物理、系统生物学、生物信息学等需要大量数据支撑的数据密集型科学。英国在科学数据管理、保存和共享方面拥有很多实践经验。

1 英国科学数据管理政策及要点

1.1 科学数据管理政策体系

英国内阁办公室 2012 年发布《开放数据白皮书》[1]，明确将数据列为国家基础设施的重要组成部分，采取支持建立开放数据研究所等措施加大政府部门公共数据开放。从 2010 年开始建设的政府数据开放平台 data.gov.uk 网站，目前有超过 4 万个政府数据集开放。英国科技管理行政部门——原商业、创新和技能部（现为商业、能源和产业战略部）2012 年以来两次发布《开放数据战略》[2]，落实该部门的数据开放共享工作。2017 年，英国议会通过《数字经济法》[3]，推动数字政府建设和政府数据共享。政府的这些规制引导和加强了数据开放共享的理念和做法。

科学数据管理政策主要由研究理事会等科研资助机构制定。英国共有 7 个研究理事会，分别是工程与物理科学研究理事会（EPSRC）、医学研究理事会（MRC）、生物技术与生物科学研究理事会（BBSRC）、自然环境研究理事会（NERC）、科技设施理事会（STFC）、经济与社会研究理事会（ESRC）、艺术和人文研究理事会（AHRC）。英国研究理事会（RCUK，现为英国研究与创新总署 UKRI）统管上述 7 个研究理事会，共发布了 3 项数据管理政策。一是 2011 年发布、2015 年修订的《研究数据管理最佳实践指南》[4]，提出了数据管理政策的七大原则；二是 2016 年 RCUK、英格兰高等教育基金会、英国大学联盟和慈善机构维康基金会公同发布的《开放研究数据协议》[5]，确定开放研究数据的十大原则，适用于公共资助的所有研究领域；三是 2013 年发布的《通过拨款资助支持研究数据管理成本》[6]，对如何在项目经费中支出研究数据管理费用做了明确规定。目前 UKRI 延续使用上述 3 项管理政策。

7 个研究理事会和维康基金会等其他重要资助

第一作者简介：王静（1977—），女，理学硕士，主要研究方向为科技创新政策和科技管理。
收稿日期：2018-05-10

机构均制定了各自的数据政策，并不断更新完善。与自然科学相关的 5 个研究理事会中，支持环境和地学领域的自然环境研究理事会于 1996 年发布了数据政策，相关规定最为完善；其他 4 个理事会在 2007—2011 年期间制定。此外，支持医学研究的医学研究理事会专门制定了针对“人群和患者研究数据共享的政策指导”[7]。

1.2　主要政策内容

研究理事会数据政策的原则和要点基本一致，主要包括以下几方面：

（1）公共资金资助的研究数据属于公共产品，为公共利益服务，应最大限度地开放。科研数据的系统管理和共享是促进高质量研究和创新的前提。

（2）各利益相关方在促进研究数据管理和共享方面应各司其职，明确责任和义务。研究机构应尽可能创造数据开放共享的环境，制定数据管理政策和程序，确保研究人员获得适当的培训和资源，保证研究人员遵守相关法律、政策和程序。研究人员应在一定时期内（具体要求因学科而异）尽可能将科研数据公开，且保证提供的数据符合最高质量标准，以便其他研究人员获取、理解和使用。资助机构应制定科研数据共享政策，并对相关做法提供经费支持。

（3）尊重法律、伦理要求和商业利益需要，必要时对开放予以限制，平衡数据开放和信息保护之间的关系。

（4）研究数据创造者对数据拥有合理的优先使用权，可将数据延迟发布。

（5）应公开充分的元数据，并对数据进行综合处理，使之便于使用和长期保存；数据应可发现，可理解，可获取。

（6）数据使用者应注明来源，遵守数据获取条款。

（7）应考虑科学数据开放共享的成本效益，建立高效、经济的机制来保证科学数据共享利益的最大化。

（8）科研机构的数据管理政策和具体项目的数据管理计划应遵循相关标准和科研惯例。

（9）对数据管理政策执行情况进行监督。多数研究理事会会监督数据管理政策的执行情况，特别是在项目结题时评估数据管理计划的实施情况。

2　英国科学数据中心布局及建设情况

2.1　重要的科学数据中心

根据英国大学联盟 2017 年发布《英国研究数据基础设施》[8] 的报告，英国目前有近 250 个研究数据库。各研究理事会的网站上列出了本领域常用的数据中心、数据资源目录和网站链接，便于研究人员使用。英国重要的数据中心整理如下。

（1）地球与环境科学领域的数据中心：自然环境研究理事会建设有 5 个数据中心，分别是英国海洋资料中心、环境数据分析中心（其中包括英国大气数据中心、自然环境研究理事会地球观测数据中心、英国太阳能系统数据中心）、环境信息数据中心（陆地和淡水）、国家地球科学数据中心和极地数据中心（极地和冰冻圈）。自然环境研究理事会要求其资助的研究项目在 2 年内将数据交至相应的数据中心；数据中心负责维护数据，并将其提供给来自社会各界的所有用户。有些中心还保存有自然环境研究理事会研究项目期间收集的实物标本和样本材料，以及第三方提供的材料。

（2）生物、医学和农学领域的数据中心：由生物技术与生物科学研究理事会、医学研究理事会、卫生部、维康基金会等支持建设。主要有：欧洲生物信息研究所（EMBL-EBI），管理和维护着世界上最全面的分子生物学数据库，向全球科技界免费提供；英格兰基因组学数据中心（Genomics England），是英国开展 10 万人类基因组测序计划的主要支撑平台；英国生物样本库（UK Biobank），收集了英国 50 万份 40~69 岁志愿者的 DNA 样本，目标是建成世界上最大的有关致病或预防疾病的基因和环境因子的信息资源库；洛桑样本库，是全世界农业研究的经典数据库，拥有来自经典田间实验的 30 万份作物、肥料和土壤样本，并仍在不断增加样本。诺丁汉拟南芥储存中心，向国际拟南芥基因组计划和更广泛的研究界提供拟南芥种子和信息资源。

（3）物质科学领域数据中心：工程与物理科学研究理事会资助的国家化学数据库服务（CDS）提供了最先进的化学数据库和工具，可以从任何英国学术网络免费访问。剑桥晶体学数据中心

（CCDC）是国际上最重要的晶体学数据中心，收集并提供具有 C—H 键的所有晶体结构。

2.2 科学数据中心的管理

（1）关于数据库选择和数据保存期限。科学数据一般通过已有数据库进行管理，研究人员应选择最可能从数据集成中获取最大科研价值的数据库，并在项目制定数据管理计划时予以明确。项目不同阶段的数据有时会储存在不同的数据库，例如，原始数据可能会存于实验设施数据库，而派生或分析数据则存于某一领域的专业数据库。一般科研数据通常至少保存 10 年，如未来可能涉及法律责任的数据；有些研究领域的特殊性质决定了需要保存更长时间，如临床研究数据要保存 20 年。

（2）关于数据汇交规范。一般应根据拟提交数据库的要求或根据学科约定俗成的惯例而提交。生物技术与生物科学研究理事会规定了需要数据共享的 3 个领域是大量实验产生的数据、长时间序列或累积方法产生的低吞吐量数据、使用系统方法生成的生物建模数据。自然环境研究理事会对数据汇交有较全面的管理规范，包括《自然环境研究理事会数据政策》和《数据政策指导说明》[9]，其中对汇交格式进行了详细要求；还制定了《关于保存自然环境研究理事会模型代码和模型输出的指导》，要求研究人员保存受自然环境研究理事会资助的研究产生的模型代码和模型数据，以便后续研究人员使用。

（3）关于汇交数据的开放共享。绝大多数数据中心的数据对研究人员免费开放，但基于数据加工产生的信息产品要收费。自然环境研究理事会出台了《自然环境研究理事会许可和收费政策》，规定了对在环境数据基础上加工产生的信息产品收费的办法。各数据中心对数据存储、安全防范、匿名保护、数据加密传输、数据共享程序等有很多细化规定。如 UK Biobank 对收到数据的共享申请，要经过 2 个专门委员会的审核，一个负责数据申请审核，另外一个负责督导。

3 国家财政支持的科研项目产生的科学数据管理

3.1 项目要制定完善的数据管理和共享计划

研究理事会等各资助机构都规定，由其全额或部分资助的项目，在提交正式项目申请时必须制定完善的数据管理和共享计划，且作为同行评议的内容之一。数据管理计划应涵盖科研项目可能产生的所有数据，包括原始数据和发布数据。数据管理计划应包括以下内容[10]：项目的数据来源；分析现在可能利用的数据与项目研究所需要的数据存在的差距；研究项目将产生的数据相关信息，包括数据量、数据类型、数据质量、数据格式、数据标准、元数据标准、数据收集方法等；数据质量保证及数据备份计划；数据共享所预期的困难及应采取的措施；数据保密性与数据使用伦理；数据知识产权；项目组成员的数据管理职责等。

3.2 评估数据是否有保存价值的标准和元数据

评估数据或数据集价值的标准有利于指导研究人员选择有价值的数据进行保存与管理。关于数据保存和公开的决定，一般应由研究人员在适当机构层面监督程序下做出，监督程序应该是可验证、透明的。研究人员应对科研数据的利用价值进行评估，以决定弃留。例如，本质上无法重新测量或再现的数据，如地球观测数据或人群数据，往往需要永久保存。但有些情况下，数据保存无法实现或不经济。这取决于数据的类型和规模、数据在验证发布结果中的作用大小、对未来研究的价值等。例如，就模拟数据或模型输出而言，保存模拟方法或模型远比保存数据本身更为有效。如不会对研究成果的验证造成根本性影响，可适时将数据弃置。

元数据指科学数据的相关信息，如项目题名、作者等信息，这是用户查找与使用数据所依赖的。为保障数据的正确利用，一般均要求数据附加相应的元数据。

3.3 关于数据公开的时滞期

由于充分肯定数据创造人员的优先使用权，以及考虑到数据公开需要一定的数据准备时间，一般都允许数据公开有一定的时滞。

数据公布的时间节点应根据数据本身的性质和种类、采用的研究方法和学科领域，通常不应晚于主要研究发现发布之日。如果资助方有特殊要求，或某一学科领域已有惯例，科研人员应予以遵守。对于支撑论文的数据，通常公布日应不晚于论文的发表日。自然环境研究理事会要求研究人员在数据采集完成后 2 年内存储至其数据中心；工程与物理

科学研究理事会要求科研人员在数据产生的 1 年内发布；生物技术与生物科学研究理事会则要求按已有惯例提交，没有惯例的在生成数据集的 3 年内发布。有些数据的产生需要经年累月的巨大努力，这种情况下可以考虑授予研究人员对数据的长期专有使用权。

数据管理计划中应配备充足的资源、在合理时间内进行数据准备工作（指补充元数据、对数据进行注释等），不能以此为借口，无限期推迟数据的发布时间。对于有商业利用价值的科研成果，可以采取适当延后数据共享的方法予以保护，但科研成果的商业化并不是数据不共享，不应以此为借口将数据共享过度延后。如研究时间跨度较长，数据可以分批发布；这种情况下应注意防止过早共享影响后续研究，注重数据的长期价值。

3.4　公开的数据要可发现、可理解、可获取

发布的科学数据应易于被其他研究人员发现和有效利用，要遵循以下三大原则。

（1）可发现：科学数据应辅以充分的背景信息，让研究人员了解哪些数据存在、如何产生和如何获取，并确定一般性或最低元数据标准。

（2）可理解：提供充分的原始数据，如描述数据来源、处理、分析和管理过程，最大限度地减少误用、误解或困惑。

（3）可获取：非数字形式的数据应注明如何获取，如因法律、伦理、商业或安全考虑须限制共享，应充分说明理由，并说明在何种条件下可以获取该数据。

3.5　对数据使用者的要求

任何数据使用者都应尊重数据创造者的劳动，在使用数据时注明出处。这既是对他人贡献的尊重，也能使数据更容易被找到，促进科研成果的复制和再现，还可以对特定研究数据的影响进行跟踪。数据引用可注明数据创造、保存和共享者，对于多方来源数据可注明数据库，应包含永久识别码等充分信息，便于找到数据确切版本。

4　科学数据保密与安全管理制度

4.1　数据保密与安全相关法律制度

科学数据保密与安全管理相关制度主要是《信息自由法》《数据保护法》等，研究理事会的数据政策需符合这些法律，科研机构和研究人员也要了解并遵从法律。

英国《信息公开法》[11] 于 2000 年 11 月通过，2005 年生效。该法规定了 20 多种信息公开的例外情况，主要包括 3 个方面 [12]：一是涉及情报安全机构的信息；二是涉及公众利益的信息；三是涉及个人的信息。

英国 2017 年发布新的《数据保护法》[13]，于 2018 年 5 月生效。主要特点是对不法行为实施更为严厉的处罚。如将未经数据拥有者同意、故意或过失使“已去识别化”的个人数据被重新识别出来的行为，增加为刑事犯罪；赋予负责数据保护法监管和执行的信息专员更多权力，对数据控制和处理人员的严重违法行为，可处以高达 1 700 万英镑或全球营业额 4% 的行政处罚；针对数据控制和处理人员非法更改记录的行为，信息官员可对其提起刑事诉讼。此外，该法对一般性数据保护设立了新的标准，赋予公民对个人数据更多的掌控权。

英国内阁办公室制定《政府安全分类》[14]，于 2014 年 4 月生效，描述了政府对信息和数据的分类。适用于政府收集、存储、处理、生成或共享以提供服务和开展业务的所有信息，包括从外部合作伙伴处收到或与之交换的信息。信息资产可分为 3 种类型：官方（Offical），秘密（Secret）和绝密（Top Secret）。政府部门及其机构应实施这一政策，因此研究理事会（属于非政府部门公共机构）应予实施。

医学研究理事会对其处理信息的标记政策，按政府安全分类系统分为官方、官方–敏感、秘密、绝密 4 类 [15]。除非另有标注，否则在医学研究理事会内接收、创建、处理、生成、存储或共享的大多数信息归类为“官方”，但并不需要标记；特别敏感的信息标记为“官方–敏感”，此标志仅在有限情况下使用；秘密和绝密适用于需要加强防护措施的高度敏感信息，但在工作过程中很少涉及。

4.2　合法合规、负责任的科学数据共享

在科学数据尽量广泛开放共享的大原则下，并非所有数据均可以无条件对外开放。出于保密、隐私保护、知识产权保护、国家安全、成本及遵循数据许可协议条款等考虑，综合考虑法律、监管和伦理要求，包括适用的数据保护、科研伦理、科研诚

信规则等，确定数据是否可以或应该开放及开放范围。另一方面，不能不加辨别地对所有研究数据开放进行限制，如要限制，须提出合理正当的理由，平衡开放和敏感信息保护的关系。

有些研究数据的不当或过早公开不仅可能影响研究工作，还会危害个人或公共利益。个人信息、敏感信息和保密信息在遵循了如隐私权、知情同意权等法律和监管要求、职业标准的前提下才可共享。如果研究要将不同数据集进行整合，而整合只能通过使用个人身份验证，则需要在可靠的环境下进行，保证个人身份识别风险降至最低。任何情况下，如涉及个人信息，在数据发布或给与第三方进行分析之前都要签署数据共享协议，严禁使用发布的数据来识别参与者身份或违反保密规定，严禁未经许可与参与者联系或解除。

还有其他不符合公共利益的数据发布情况也应予以限制，例如能识别受保护物种位置的数据等。

5 科技合作产生的科学数据的管理

5.1 公共资金资助的研究项目产生的科学数据管理主要用合作协议来约束

研究理事会数据政策规定，对于公共资金资助的产学合作研究项目，企业处于商业考虑，可能对其提供的数据限制共享。即使在这种情况下，合作研究所发表的论文也应明确如何及在何种条件下可以获取支撑数据及其他研究资料。合作协议中应事先明确谁有权获取这些数据，同时，学术界科研人员在提交拟发表论文之前，应获得企业合作方的同意。如事先未签署上述合作协议，研究理事会希望企业合作方对科研成果的公布不受限制。

与非英国机构的数据共享应遵循同样的原则和标准。采取所有合理措施，避免将研究数据保存在法律保护程度低于英国的国家或地区。科研人员应确保合作研究中产生的知识产权得到适当保护和管理；如果已取得具有潜在商业价值的知识产权，在有必要推迟数据共享的情况下，可留有一定合理时间采取知识产权保护措施，如起草和申请专利或进行许可申请等，但应最大限度地减少这种由于知识产权管理而进行的共享推迟或限制。

5.2 企业自行研发投入产生的科学数据的管理

主要约束法律为《数据保护法》，要确保研发过程中涉及到的个人数据隐私和安全。此外该法规定："企业收集的所有用户数据必须能够以标准化格式供用户下载。"这将极大促进企业在数据格式标准化方面的投入。

此外对于制药企业，在研发过程中要遵循美国食品药品监督管理局（FDA）、欧盟欧洲药品管理局（EMA）等要求的文件资料管理规范，如保留所有旧版和各版本修改记录，支持电子签名、修改留痕等，确保材料齐全并随时做好封存备查准备。

6 针对我国科学数据管理工作的建议

6.1 制定操作性强的实施方案，加大数据共享办法的落实

2018 年 3 月国务院发布了《科学数据管理办法》[16]，是我国科学数据管理的里程碑事件。建议有关部门组织制定更加具体的实施办法，推动落实。如可发挥专业学会等专业共同体的作用，对某学科领域哪些数据需要保存和共享予以明确，并研究制定数据标准、元数据、汇交时滞等规则；要求各法人单位制定操作性强的数据管理办法，从而加强对科学数据的最初创造者和最终使用者——研究人员的指导，营造既鼓励开放共享又充分尊重数据创造者优先使用权的政策环境等。

6.2 科学数据中心的建设要加强统筹、强调专业性，不宜重复分散

从国家层面加强科学数据的集中汇交和管理，建设数据全、质量高、开放共享的国家数据中心非常重要。为避免条块分割、资源分散或重复建设，建议有关部门加强统筹，发挥科技项目管理专业机构、专业学会等的作用，在现有资源聚集、优势突出的数据中心基础上择优支持一批按专业领域分类的国家级数据中心，制定相应的数据汇交、管理和共享政策，并予以长期支持。

6.3 对数据跨境管理予以研究和规制

信息社会下，信息和数据的流动常常是跨越国界的，难以监管。如何既保护国家秘密数据安全与个人隐私，又最大限度地发挥数据作用，是各国都面临的现实问题。英国近年出台《数字经济法》《数据保护法》《数据伦理框架》等规章制度，建设数据伦理和创新中心等，在数据管理和共享的规则制定方面有一定的探索，我国可以借鉴，加强对科学

数据跨境流动的研究和规制。■

参考文献：

[1] Cabinet Office. Open data white paper[EB/OL]. [2018-04-29]. https://www.gov.uk/government/publications/open-data-white-paper-unleashing-the-potential.

[2] BIS. BIS open data strategy 2014 to 2016[EB/OL]. [2018-04-29]. https://www.gov.uk/government/publications/bis-open-data-strategy-2014-to-2016.

[3] UK Parliament. Digital Economy Act 2017[Z/OL]. [2018-04-29]. https://www.gov.uk/government/collections/digital-economy-bill-2016.

[4] RCUK. Guidance on best practice in the management of research data[EB/OL]. [2018-04-29]. https://www.ukri.org/files/legacy/documents/rcukcommonprinciplesondatapolicy-pdf/.

[5] RCUK. Concordat on open research data[EB/OL]. [2018-04-29]. https://www.ukri.org/files/legacy/documents/concordatonopenresearchdata-pdf/.

[6] RCUK. Supporting research data management costs through grant funding[EB/OL]. [2018-04-29]. https://blogs.rcuk.ac.uk/2013/07/09/supporting-research-data-management-costs-through-grant-funding/.

[7] MRC. MRC policy and guidance on sharing of research data from population and patient studies[EB/OL]. [2018-04-29]. https://mrc.ukri.org/publications/browse/mrc-policy-and-guidance-on-sharing-of-research-data-from-population-and-patient-studies/.

[8] Universities UK. Research data infrastructures in the UK[EB/OL]. [2018-04-29]. https://www.universitiesuk.ac.uk/policy-and-analysis/research-policy/open-science/Documents/ORDTF%20report%20nr%201%20final%2030%2006%202017.pdf.

[9] NERC. NERC data policy[EB/OL]. [2018-04-29]. https://nerc.ukri.org/research/sites/data/policy/.

[10] 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.

[11] UK Parliament. Freedom of Information Act[Z/OL]. [2018-04-29]. https://www.legislation.gov.uk/ukpga/2000/36/contents.

[12] 夏镇平 , 高抒宇 . 英国中央政府政务公开的主要做法和经验 [J]. 中国行政管理，2006，247（1）：91-94.

[13] UK Parliament. Data Protect Act[Z/OL]. [2018-04-29]. https://www.gov.uk/government/collections/data-protection-act-2018.

[14] Cabinet Office. Government security classifications[EB/OL]. [2018-04-29]. https://www.gov.uk/government/publications/government-security-classifications.

[15] MRC. MRC document marking policy[EB/OL]. [2018-04-29]. https://mrc.ukri.org/about/information-standards/document-marking-policy/.

[16] 国务院办公厅 . 科学数据管理办法 [EB/OL]. [2018-04-29]. http://www.gov.cn/zhengce/content/2018-04/02/content_5279272.htm.

An Overview on the Management of Research Data in the UK

WANG Jing[1], MA Hui-qin[2]

(1. Ministry of Science and Technology of the People's Republic of China, Beijing　100862;
2. China Science and Technology Exchange Center, Beijing　100045)

Abstract: The open sharing of research data has become the focus of open access in recent years. The UK is the first country to propose the e-science concept and has a lot of practical experience in research data management and sharing. The research data management in UK is outlined in this paper, including management and sharing policies, the construction of research data centers, the management and sharing mechanisms of research data generated by public finance-supported research projects, the research data confidentiality and security management. Relevant suggestions for research data management in China are also proposed.

Key words: UK; research data; data management; data sharing