现代情报
2008年11月 第11期
November 2008 No. 11

国际科学数据共享研究

江 洪[1,2] 钟永恒[2]
(1. 武汉大学信息管理学院，湖北 武汉 430072；
2. 中国科学院国家科学图书馆武汉分馆，湖北 武汉 430071)

〔摘 要〕 科学数据作为信息时代最基本、最活跃、影响面最宽的科技资源和一种战略性资源，对于科技创新具有显著的基础支撑作用。本文描述了科学数据的特征，以美国、英国和欧盟科学发展计划为主，研究了欧美国家科学数据共享发展状况；以国际科学理事会、经济合作与发展组织等国际组织为例，研究了一些国际组织推动科学数据共享的发展计划，介绍了世界著名科学数据中心的科学数据共享情况，以此对科学数据共享发展的过程进行了初步探究。

〔关键词〕 科学数据共享；欧美国家；国际组织；数据中心
〔中图分类号〕G250.74 〔文献标识码〕A 〔文章编号〕1008-0821 (2008) 11-0056-03

Study of the International Scientific Data Sharing

Jiang Hong[1,2] Zhong Yongheng[2]
(1. School of Information Management, Wuhan University, Wuhan 430072, China;
2. Wuhan Branch, National Science Library, CAS, Wuhan 430071, China)

〔Abstract〕 The sciences data as a kind of the scientific resource and strategic resource has the most foundational and active characters. It is the important basic support to the innovation of sciences. In this paper, the foundational character of scientific data was stated. According to research some scientific data sharing plans in some countries and international organizations, such as USA, EU, International Council for Science, and so on, the status of the development scientific data sharing was analyzed.

〔Key words〕 scientific data sharing; USA and EU; international organization; data center

科学数据是人类社会从事科技活动所产生的原始观测数据、探测数据、试验数据、实验数据、调查数据、考察数据、遥感数据、统计数据、研究数据以及相关的元数据和按照某种需求系统加工的数据，具有科学价值和使用价值。科学数据是信息时代最基本、最活跃、影响面最宽的科技资源和一种战略性资源，它对于科技创新具有显著的基础支撑作用[1]。

1 科学数据特征

科学数据是具有明显潜在价值并可在广泛应用中得以增值的巨大社会财富。显然，科学数据的价值决定于社会对科学数据使用的需求，科学数据是一种特殊的社会资源，除了具有一般资源的价值属性外，还有着与其它资源不同的特征。

(1) 科学数据需要长期积累，其准确性系统性是科学数据所具备的基本特点。人类科技活动产生并长期积累着科技数据，它又可以按照社会的多种需求提供或可能提供系统的、足够的数据。

(2) 科学数据是需要保证不断的更新和补充，不同的研究人员产生的研究数据通过可共享的数据管理，使得数据更加完善，也更加可靠。

(3) 科学数据数量巨大，海量数据通过一些人为调控后，其提供的质量、产品形态及其存储与传输方式都会更利于使用，而且，其价值的实现与开发者的能力和方法密切相关。

(4) 科学数据具有更为特殊的非排它性，即科学数据的应用不限于本专业、本领域，可为不同的研究者，从不同的角度去挖掘各自所需的科学技术、社会、经济等不同的知识。

(5) 数据可以无限制复制的特点，决定了它不会因为满足某人、某时的需求而影响他人对其的需求。科学数据这些共享特性的充分发挥，将使其效用价值倍增。

不难看出，科学数据应该共享，其共享的逻辑起点是它的资源属性。科学数据只有在广泛应用过程中，才能实现数据拥有者对其获取最大效用的追求。同时，又可在应用过程中衍生出满足更高层次需求的新数据，使得这一社会广泛需求的公共物品面向全社会共享的行为变得复杂[4]。建立健全符合客观规律的共享机制，会使科学数据的这种特殊社会资源得到高效、有序的管理，更加合理地更大效率地发挥作用，从而促进社会进步与发展。

收稿日期：2008-07-31
作者简介：江 洪 (1968-)，女，研究员，研究方向：情报学，发表论文20多篇。
钟永恒 (1965-)，男，研究员，研究方向：情报学，发表论文数篇。

2 欧美国家的科学数据共享

国外的科学数据共享发展于上世纪 70 年代，美国可以说是科学数据共享的先导者。在上世纪 70 年代美国的科学数据积累迅速增加。据估算，数据库总量一直占据全世界总量的一半以上。到 1975 年，美国开发了 177 个大型数据库，主要服务目标是政府决策和政府启动的重大科研项目[5]。

20 世纪 90 年代初美国将“完全与开放”的数据共享政策作为美国联邦政府在信息时代的一项基本国策，通过数据的流动和应用激励美国经济的发展，确保美国在 21 世纪信息时代科技和综合国力处于世界领先地位[6]。

2006 年 7 月美国科学基金会发表了“21 世纪科学信息化基础架构（NSF S cyberinfrastructure Vision for 21st Century Discovery）[7]”的报告，报告指出“在未来，美国科学和工程上的国际领先地位将越来越取决于在数字化科学数据的优势上，取决于通过成熟的数据挖掘、集成、分析和可视化工具将其转换为信息和知识的能力。”因此，他们在国内机构层面、国家层面及国际层面上采取一系列的行动计划，目的是准备建立全球性的科学研究数据的公共访问机制，为美国科学研究和经济建设服务，并保持美国在科学研究和工程技术领域的领先地位。

2000 年，英国政府提出了《e - Science 计划》[8]，总经费 2.5 亿英镑。建立了 e - Science 中心，随后又扩展了一批专业优势中心。同时还支持了一批 e - Science 项目；开发通用网格中间件；促进技术辐射和国际合作。e - Science 的第二阶段为 2003 - 2006 年。计划继续支持已有的 e - Science 中心。为了加强数据方面的工作，还成立了一个全国数据管理中心。

2004 年，英国财政部、工贸部和教育部发布了《科研与创新投资框架》（Science and Innovation Investment Framework 2004 - 2014）[9]。2007 年 3 月，发布了研究报告《发展英国科研与创新信息化基础设施》（Developing the UK's e - infrastructure for science and innovation）[10]。提出数据资源数字化长期保存与共享建设规划，重点要建立大规模的国家科学数据中心。报告指出：英国知识经济的增长十分依赖于对研究人员的持续支持，特别是他们和产业界的合作，以及世界领先的创新成果的商业化应用。国家科研信息化基础设施（e - infrastructure for research）为英国科学研究提供了一个至为关键的基础平台，不仅支持了技术的迅速发展，而且也为转移知识和创造财富提供了新的可能。报告重点强调了科学数据建设国家科研信息化基础设施的关键问题，包括：科学数据和信息的产生，数据的保存和管理，数据的查询和导航，虚拟研究团体，网络、计算和数据存储设施，AAA（认证 authentication、授权 authorization 和核算 accounting），中间件（middleware）和数字版权管理。重点关注科学数据如何为更广泛的科研人员和产业服务的问题。

欧盟在 2002 年，发表了《迈向信息社会：原则、战略和优先行动》布加勒斯特宣言（The Bucharest Declaration），提出对公共科学数据、公共当局持有的信息开放共享的公益性共享原则和指导思想[11]。

欧盟信息基础设施咨询工作组 e - IRG（e - Infrastructure Reflection Group），成立于 2003 年 12 月，目前由 29 个欧洲国家的国家代表和欧盟代表组成，其任务是在政策、咨询和监督的层面，包括技术和管理问题，提出相关政策和管理模式的建议，以便在欧洲范围内经济和方便地共享信息化资源[12]。

3 国际组织的科学数据共享

在科学数据共享领域，国际组织也发挥了重要的作用，科学数据共享活动的顺利长期的开展，需要世界范围内的国家、组织、机构间的合作与交流。国际组织通过相关努力推动全球范围内的科学数据广泛应用。

国际科学理事会 ICSU 是目前科学界权威的非政府科学组织，其下属的科技数据委员会 CODATA 和世界数据中心 WDC 是国际研究科学数据管理和应用的专门组织。为了支持对研究和教育数据的“完全与开放”获取，CODATA 在 2000 年制定了《网络时代的科学原则》[13]。

2002 年 CODATA 针对发展中国家的数据开发和利用专门成立了“发展中国家数据保护与共享任务组”[14]（The CODATA Task Group on Preservation of and Access to Scientific and Technical Data in Developing Countries），促进世界范围内更深入的理解发展中国家对科技数据的长期保存、归档管理和共享等活动中遇到的困难和必要的条件。

世界数据中心（World Data Center）是国际科学联合会下设的科学数据组织，目前有 40 多个学科数据中心，分属 4 个数据中心群：WDC - A 美国、WDC - B 前苏联、WDC - C 欧洲和日本、WDC - D 中国。2005 年由 UNESCO、ICSU、OECD 等组织共同资助提出的主题为“建立科学信息共有，面向机构政策和行动指南”的“全球科学信息共有先导计划”[15]激励人们尝试新模式的创造、传播和合作利用科学数据信息。

经济合作与发展组织 OECD 为了帮助成员国内部公共机构、研究者对科学数据的收集和公共获取，于 2006 年颁布了《公共资金资助的研究数据获取原则与指南》[16]，提出 13 条原则指导成员国制定并完善科学数据共享政策，形成了国家间的共享共识。该《原则与指南》要求成员国在进行公共领域的科学数据共享活动时，将这些原则用在制定国家法律和研究政策中。

2007 年，由联合国教科文组织（UNESCO）批准、中国科学院等单位领衔的“促进发展中国家科学数据共享与应用全球联盟（UN e - SDDC）”计划（Global Alliance for Enhancing Access to and Application of Scientific Data in Developing Countries）[17]正式启动，力图在科技界可持续发展方面填补数字鸿沟，缩小发达国家与发展中国家之间的数字差距。

4 著名科学数据中心的数据共享

4.1 美国航空航天局(NASA)分布式最活跃数据档案中心群(DAACs)[18]

NASA 与科学数据相关的机构主要是空间科学数据运行办公室，该机构下设有美国国家空间科学数据中心（NSSDC）和空间物理学数据运行中心（SPDF），数据资源集中在天文和空间科学领域，数据主要来自于 NASA 的空间飞行计划。NSSDC 负责 NASA 数据永久存档，提供天体物理学、空间物理学等数据。SPDF 主要负责多任务和多学科的数据服务的设计和实现。

1990 年，美国航空航天局（NASA）着手建设分布式最活跃数据档案中心群（DAACs—Distributed Active Archive Centers），由此标志着美国国家层面上的科学数据共享工作划时代的开始。根据美国新一代地球观测系统计划，美国

DAACs的数据存储量将由1998年的0.8PB增加到2010年的18PB，这将使美国在21世纪继续在科学成就和国家实力方面保持世界领先地位。

4.2 美国国家生物技术信息中心(NCBI)[19]

NCBI是在NIH的美国国家医学图书馆（NLM）的一个分支。美国国家医学图书馆（NLM）因为在创立和维护生物信息学数据库方面的经验而承担了建立一个内部的关于计算分子生物学的研究计划。这也是在科学数据共享领域专业图书馆所具优势的最好体现。而且，美国的科学数据共享体系中，专业图书馆被作为一类重要的数据中心而建设。

NCBI的任务是发展新的信息学技术来帮助对那些控制健康和疾病的基本分子和遗传过程的理解。它的使命包括4项任务：

（1）建立关于分子生物学，生物化学，遗传学知识的存储和分析的自动系统。（2）实行关于用于分析生物学重要分子和复合物的结构和功能的基于计算机的信息处理的，先进方法的研究。（3）加速生物技术研究者和医药治疗人员对数据库和软件的使用。（3）全世界范围内的生物技术信息收集的合作努力。

NCBI自1988年建立的关于分子生物学、生物化学和遗传学的数据库和数据分析系统，推动生物信息学领域数据库和数据分析软件的使用，开展计算机生物信息处理先进方法的研究。数据主要来源于两部分：美国各实验室提交的基因序列数据和同国际上的基因数据库交换的数据。

NCBI提供数据资源网络共享，主要是通过NCBI开发的一系列工具和软件实现的，如基因序列注册软件BankIt，数据搜索软件Entrez，基因序列比对分析软件BLAST等。

NCBI还通过赞助会议、研讨会、系列演讲、科学访问学者项目、提供博士后工作位置等方式来开展在应用于分子生物学和遗传学的计算机领域的科学交流。

4.3 美国国家大气研究中心(NCAR)研究数据归档中心[20]

美国国家大气研究中心始建于1960年，是大气及相关科学问题的研究中心，数据资源集中在大气科学领域。主要有大气分析格点资料、卫星资料、长年代的气候资料、海洋资料等。目前有400多个观测和分析资料的数据集，并将持续增加。其数据面向全美科学家、教师和学生，提供网络数据共享。

4.4 日本产业技术综合研究所(AIST)[21]

AIST是目前日本规模最大的科研院所，科研专家和重点主要集中在生命科学与IT技术、电子学与纳米技术、制造业与环境科学、能源与测量技术、材料科学、地球科学六大领域。科研数据公开数据库（RIO－DB）拥有70个主题数据库，数据来源于AIST各机构的科研项目，数据整理工作由AIST各研究机构完成。全部数据库通过网络提供免费服务，服务于科研机构，也服务于一般工业企业。

参 考 文 献

[1] 黄鼎成，李晓波，王卷乐. 浅谈科学数据共享工程建设的战略取向 [J]. 中国基础科学，2005，(5)：29－35.

[2] 严冬梅，尚翔. 论科技创新的基石——科学数据共享 [J]. 管理科学研究，2005，23 (1)：20－23.

[3] 魏东原，朱照宇. 专业图书馆如何实现科学数据共享 [J]. 图书馆论坛 2007，27 (6)：253－256.

[4] 黄鼎成. 科学数据共享的理论基础与共享机制 [J]. 中国基础科学，2003，(2)：22－27.

[5] 美国国有科学数据的“完全与开放”共享国策 [EB]. http://www.qiji.cn/scinews/detailed/838.html，2008－07－22.

[6] 陈传夫，曾明. 科学数据完全与公开获取政策及其借鉴意义 [J]. 图书馆论坛，2006，26 (2)：1－5.

[7] NSF Cyber infrastructure Council. 2006. NSF's cyber infrastructure vision for 21st century discovery [R]. CI DRAFT: Version 7.1 National Science Foundation USA. http://www.nsf.gov/news/special_reports/cyber/index.jsp，2008－07－18.

[8] National e－Science Centre [EB/OL]. http://www.nesc.ac.uk，2008－07－18.

[9] science and innovation investment framework 2004－2014: next steps [EB/OL]. http://www.hm－treasury.gov.uk./media/7/8/bud06_science_332v1.pdf，2008－07－18.

[10] Developing the UK's e－infrastructure for science and innovation [EB/OL]. http://www.nesc.ac.uk/documents/OSI/index.html，2008－07－18.

[11] The Bucharest Declaration，Bucharest Pan—European Conference in Preparation of the World Summit on the Information Society: Towards an Information Society: Principles，Strategy and Priorities for Action [R]. Bucharest: 9 November 2002.

[12] e－Infrastructures: a fundamental building block of the ERA [EB/OL]. ftp://ftp.cordis.lu/pub/ist/docs/rn/karlson.pdf，2008－07－18.

[13] Principles for science in the internet era [EB/OL]. http://www.codata.org/dataaccess/principles.html，2008－04－22.

[14] The CODATA Task Group on Preservation of and Access to Scientific and Technical Data in Developing Countries [EB/OL]. http://www.tgdc－codata.org.cn/，2008－04－22.

[15] Creating the Information Commons for e－Science: Toward Institutional Policies and Guidelines for Action [EB/OL]. http://www.codataweb.org/UNESCOmtg/index.html，2008－04－22.

[16] Recommendation of the Council concerning Access to Research Data from Public Funding [EB/OL]. http://webdomino1.oecd.org/horizontal/oecdacts.nsf/Display/3A5FB1397B5ADFB7C12572980053C9D3?OpenDocument，2008－04－22.

[17] Global Alliance for Enhancing Access to and Application of Scientific Data in Developing Countries [EB/OL]. http://www.e－sddc.org/cn/index.html，2008－04－22.

[18] Distributed Active Archive Centers [EB/OL]. http://nasadaacs.eos.nasa.gov/about.html，2008－07－18.

[19] National Center for Biotechnology Information [EB/OL]. http://www.ncbi.nlm.nih.gov/，2008－07－18.

[20] the National Center for Atmospheric Research [EB/OL]. http://www.ncar.ucar.edu/，2008－07－18.

[21] National Institute of Advanced Industrial Science and Technology [EB/OL]. http://www.aist.go.jp/index_en.html，2008－07－18.