

国外开放科学数据研究综述*

黄永文 张建勇 黄金霞 王 昉

(中国科学院国家科学图书馆 北京 100190)

【摘要】介绍开放科学数据的含义和开放科学数据的相关政策,分析科学数据的引用格式及 3 个主要的开放科学数据仓储,总结开放科学数据的 4 种发现服务,提出图书馆可以开展科学数据的培训服务、存储和管理服务,以及开放科学数据的发现和再利用服务。

【关键词】开放科学数据 开放数据再利用 数据引用 数据仓储

【分类号】G250

Research on the Open Research Data

Huang Yongwen Zhang Jianyong Huang Jinxia Wang Fang

(National Science Library, Chinese Academy of Sciences, Beijing 100190, China)

【Abstract】The paper introduces the definition and policy of open research data, analyzes research data citation and three major data repositories, and summaries four types discovery service of open research data. Finally, it presents library could work on training service, storage and management service of research data, and discovery and reuse service of open research data.

【Keywords】Open research data Open data reuse Data citation Data repository

1 引言

开放科学数据起源于科学研究和学术创新的要求,推动科学数据开放共享的动力主要来自国家政府机构、基金会、学会协会和出版者。相关的数据存档与数据共享的规定和政策,极大地促进了科研数据的开放访问,使科学家可以对公布的调查结果进行验证,探索新的分析方法。科研数据共享会营造更丰富的数据环境,使研究人员能够轻松地发布、发现、访问和使用数据,科研数据开放获取的基础环境正在形成^[1]。

在过去的 10 年中,欧洲和美国科学数据基础设施的发展已经取得了显著效果,并开始着手全球范围的数据共享和协作,如 OpenAIREplus(2nd Generation of Open Access Infrastructure for Research in Europe, 欧洲第二代科研开放获取基础设施)^[2]、iCORDI(International Collaboration on Research Data Infrastructure, 科研数据基础设施的国际协作)^[3],试图把开放获取的资源范围从文献出版物扩大到科学数据集,构建一个全球性的科研数据基础设施,提供开放的、可扩展的、可靠的数字科学的“生态系统”,实现全球研究数据仓储的关联和协作。欧洲研究大学联盟(The League of European Research Universities, LERU)于 2012 年 12 月发表开放科研数据的声明^[4],指出开放科研数据代表科学研究方式的革命性转变,需要重视向数据驱动的研究转变,大学、研究人员和研究资助者应共同努力,共享基础设施和最佳实践。

收稿日期:2013-04-16

收修改稿日期:2013-05-10

* 本文系中国科学院国家科学图书馆项目“开放科技资源获取与利用的技术规范研究”(项目编号:馆 1214)的研究成果之一。

本文主要从开放科学数据的含义、相关政策、数据注册和引用、数据仓储以及开放科学数据的发现角度出发,对国外开放科学数据的发展进行系统的梳理,以期对国内开展开放科学数据研究起到抛砖引玉的作用。

2 开放科学数据的含义

开放科学数据的英文形式主要有“Open Scientific Data”、“Open Research Data”、“Open Data”、“Open Access Data”。

维基对“开放数据”的定义如下^[5]:数据可以被任何人自由获取和再发布,而没有来自版权、专利或其他机制的限制。开放数据运动与其他的一些开放运动有类似的目标和宗旨,例如开源、开放内容及开放获取运动。

开放知识基金会(Open Knowledge Foundation, OKF)描述了数据开放性^[6]:内容或数据是开放的,可以免费使用、重用,并重新发布,大多数仅要求以署名-相同方式共享协议再分发。科学数据包括以支撑或附加信息等形式与论文同时出版的数据、与实验相关的图表中的数据、捕捉和报告科学现象的图像(如蛋白质凝胶)以及采集原始资料的音视频数据。

欧洲研究大学联盟(LERU)^[4]认为开放科学数据是开放知识的一部分,开放知识可以是任何类型的信息,如基因信息、地理信息、统计信息等,可以被自由使用、重复使用和重新发布。

美国科技政策办公室(OSTP)定义了开放数据的范围^[7],包括科学界普遍接受的数字记录的事实材料、需要验证的研究成果,包括支持学术论文而使用的数据集,但不包括实验笔记、初步分析、科学论文的草稿、未来的研究计划、同行评议报告、与同事的交流或物理对象(如实验室标本)。

广义的开放数据是指所有数字形式的开放数据,包括图书馆中的文献书目数据等。本文研究的开放数据主要是狭义的科学数据,指通过收集、观察和创造的各种实验数据、观测数据、统计数据、仿真数据,表现为表格、数字、图像、多媒体或其他格式。可以是论文后附带的实验数据,也可以是独立的原始数据,包括对数据进行描述的元数据、数据集以及与数据相关的出版物。

3 开放科学数据的相关政策

开放科学数据不断的发展主要是由下而上的需求和由上而下的政策共同推进的。科研人员强烈要求有一种更广范围的数据公开与再使用,呼吁对科研数据的无门槛访问(如对实验数据的开放与获取等),研究基金组织、科研机构、国家政府、期刊联盟组织等开始陆续出台科学数据的管理政策和共享政策,以促进科学数据的保存和开放访问。

(1) 科研人员发表的“潘顿原则”(Panton Principles)

开放获取在科学研究及科学交流中的作用激发了科研人员对开放数据的强烈需求,科研人员提出真正的开放获取不仅仅是对全文的获取,还应该包括对数据的获取。Murray - Rust 等^[8]科学家于2010年2月发布科学数据共享的“潘顿原则”,提出开放科学数据应该能通过互联网允许任何人以任何目的免费获取,包括下载、复制、分析、重新处理、导入软件,而没有资金、法律、技术的障碍。

“潘顿原则”所指的科学数据主要是公共资金资助的科学数据,一般不涉及私人数据。为了更有效地使用数据和方便他人添加数据,“潘顿原则”建议开放数据遵循OKF关于“开放知识/数据”的定义,采用《公共领域的贡献与许可》(Open Data Commons Public Domain Dedication and Licence, PDDL)或者《知识共享的豁免》(Creative Commons Zero)来发布和出版科学数据。

(2) 国家及基金组织、科研机构的相关政策

早在2004年,经济合作与发展组织(OECD)成员国的科技部长签署了一个宣言,提倡所有的获公共资金支持得到的研究数据都应能被公众获取、共享,并于2007年发表《公共资助科学数据开放获取的原则和指南》^[9]。随后,美国、英国等一些重要基金机构也提出了数据开放政策指南,如美国国家科学基金会(NSF)^[10]、美国国立卫生研究院(NIH)^[11]、美国国家航空航天局(NASA)^[12]、英国研究理事会(RCUK)^[13]等,要求资助的项目提交科研数据的管理与共享计划,内容包括数据范围和数据类型、标准和元数据、数据共享方式、数据权限和时限等。2013年2月,美国科技政策办公室(OSTP)签署关于“提高联邦政府资助的科学研究结果的访问”的备忘录中提到^[7],由联邦资金全部资助或部

分资助,所产生的非保密的科学研究数据应该存储并为公众提供免费的最大化访问,支持数据的检索、查询和分析。

为了确保数据及时公开发布,同时保证数据拥有者的著作权,GHG - Europe 欧洲数据政策(GHG - Europe Data Policy)^[14]将数据库共享分成公共领域和内部项目领域。除非个人版权的限制或推迟出版,公共领域数据库中的数据将立即被公开。从产生数据到出版阶段,数据的验证和分析不得超过两年。GHG - Europe 鼓励采用开放数据库协议(Open Data Commons Open Database License)来发布数据,该协议是针对整个数据库而不是每个数据集,意味着用户对整个数据库享有访问的权限。

(3) 期刊出版社的相关政策

很多期刊出版社开始实施数据存档政策,目前期刊对科学数据的存档要求主要分为两种:

①建议和鼓励研究人员在发表论文时,同时提交相关的附加数据,如 PLOS One、BMC Evolutionary Biology、F1000 Research 等。PLOS One 自 2008 年就有了数据共享的政策^[15],建议将数据保存在开放获取的机构仓储或常用数据仓库(如 GenBank、Dryad)中,或作为发表论文的支持文件。BMC Evolutionary Biology 于 2009 年在论文发表的政策中规定^[16],作者需要提交文章中所描述的材料(包括所有相关的原始数据),并承诺将免费提供给科研人员用于非商业目的研究,不过并未要求对数据进行网络存储;

②作为出版的条件,要求研究人员发表论文时,必须将论文中涉及到的研究数据提交到合适的仓储中,如 Nature、BioMed Central 等。Nature^[17]要求作者将期刊论文的数据或附加信息存储于可公开访问的数据库中(如 GenBank、Protein DataBank 等),对于有些类型的数据集(如 DNA 和蛋白质序列、大分子结构等)需要同时提供数据的访问控制号。

Dryad 于 2011 年 1 月提出联合数据存档政策(Joint Data Archiving Policy, JDAP)^[18],许多主流的期刊开始认可并实施 JDAP 政策,或使用它作为政策模型,如 The American Genetic Association、Molecular Ecology、Journal of Heredity 等。JDAP 政策要求作为出版的条件,支持论文观点和结论的数据应在合适的公共仓储中进行归档。数据是科学研究的重要产品,它们应该被保留,并在未来几十年是可用的和可重复的,数据应该带有详细的描述,发表论文中的结论应该可以利用数据重新生成。作者可以选择将数据与论文同时公

开出版,或者选择论文出版后的一定时间后(一年内)再公开科研数据。

4 科学数据的开放注册与引用

欧洲科学数据的高级专家小组在 2010 年发表的报告中^[19]指出“需要开发和利用新的方法来测量数据的价值,并奖励那些贡献数据的研究人员。”数据引用(Data Citation)是衡量科研价值的重要方式之一,可以激励科研人员以可持续的方式存档和管理数据,促进科研人员开放和共享其科研数据。以前的数据主要由研究人员自己重复使用,开放数据提高了第三方对数据的引用和再利用,Piwowar 等^[20]研究得出至少在基因表达微阵列数据方面有相当一部分被第三方重用,自 2003 年以来数据集重用的强度一直在稳步增加。有些学者研究表明在天文学领域^[21]和古海洋学领域^[22],数据共享和文献与数据之间的关联提高了文章的引用率。

为了便于科学数据的发现、引用和引用跟踪,注册科学数据集和分配唯一标识符,使科学数据集能够成为独立的、可引用的、唯一的科学对象是非常必要的。目前,科研数据唯一标识形式主要有 DOI(Digital Object Identify)、ARK(Archival Resource Keys)等,其中,DOI 是全球广泛使用的一种标识方式。DataCite 目标是促进在互联网上更容易获得科研数据,增加研究数据的可引用性,截至 2013 年 3 月底 DataCite 已为研究数据注册超过 142 万个 DOI^[23]。越来越多的系统采用 DOI 作为数据引用标识,如 Dryad、PANGAEA、eScholarship(CDL)等开放数据仓储以及 GigaScience 等开放数据期刊。

科学数据的引用已经成为近几年重要的研究主题之一,国外一些组织机构和项目对数据引用进行了探索并取得一些成果。CLADDIER 项目^[24]、SageCite^[25]、DCC^[26]等探讨了如何引用数据集、以及数据引用的最佳做法。DataCite^[23]、OECD^[9]、Dryad^[27]、Dataverse Network^[28]、Gesis^[29]等提出了具体的数据引用规范,它们的数据引用格式及样例如表 1 所示。DataCite 是针对科学数据提出的通用数据引用规范,其他的则是针对各自的数据类型和特点而提出的数据引用规范。总体来说,数据引用中至少要包括以下 4 个元数据:数据名称、数据作者、发布日期和数据的唯一标识。

表1 数据仓储推荐的数据引用格式

数据仓储	数据类型	数据引文格式及样例
DataCite	科学研究数据	< Creator > (< Publication Year >) : < Title > . < Publisher > . < Identifier > 例子: Irino, T; Tada, R (2009) : Chemical and mineral compositions of sediments from ODP Site 127 - 797. Geological Institute, University of Tokyo. http://dx.doi.org/10.1594/PANGAEA.726855
OECD	OECD 数据	< Author > (< Year of Publication Date >) , " < Dataset title > : < Dataset subtitle > " , < Parent dataset collection title > (database) < doi > < doi link > (Accessed on < date >) OECD (2008) , " Social Expenditures aggregates " , OECD Social Expenditure Statistics (database) . doi : 10.1787/000530172303 http://dx.doi.org/10.1787/000530172303 (Accessed on 21 December 2008)
Dataverse	定量数据	< Author > , < date > , " < Title > " < Identifier > < UNF > 例子: Micah Altman; Karin MacDonald; Michael P. MacDonald, 2005, "Computer Use in Redistricting", hdl:1902.1/AMXGCNKCLU UNF:3:J0PkMygLPflyT1E/8x) /EA = = http://id.theidata.org/hdl%3A1902.1%2FAMXGCNKCLU
Dryad	支撑发表论文的数据	< Creator > (< Publication Year >) Data from : < Title > . Dryad Digital Repository. < Identifier > 例子: Heneghan C, Thompson M, Billingsley M, Cohen, D (2011) Data from: Medical - device recalls in the UK and the device - regulation process: retrospective review of safety notices and alerts. Dryad Digital Repository. http://dx.doi.org/10.5061/dryad.585t4
Gesis	社会科学数据	< Author > (< publication Year >) : < Study title > . < Survey period > . < Origin/distributor > . < Study Identifier > < Object type > < Version - ID > , < DOI > 例子: ISSP Research Group (2009) : International Social Survey Programme: Leisure Time and Sports - ISSP 2007. GESIS Data Archive, Cologne. ZA4850 Data file Version 2.0.0, doi:10.4232/1.10079

5 开放科学数据仓储

越来越多的大学和研究中心开始建立科学数据仓储库,越来越多的期刊开始建议或者强制作者在公共仓储中提交数据。由于学科的要求,数据存储库是非常庞杂的,对于研究人员来说选择适当的科研数据仓储库变得困难。因此,出现了一些数据仓储的注册和目录系统,如 OAD、re3data.org、Databib 等,帮助人们识别和查找科研数据的在线信息库。开放获取目录

(OAD)中收集了开放数据仓储的列表,主要是按照学科分类列出数据仓储名称、URL 和数据仓储的简单描述。re3data.org 是全球研究数据仓储库的注册系统, Databib 是数据仓储的书目描述系统。截至 2013 年 4 月 10 日,OAD 收录 111 个开放数据仓储^[30], re3data.org 收录 338 个数据仓储(其中开放数据仓储 149 个)^[31], Databib 收录 551 个数据仓储^[32]。目前,生物、化学、地理领域的开放科学数据仓储相对多一些。本文从中选出三个重要的数据仓储进行介绍和比较,如表 2 所示:

表2 三个开放科学数据仓储的比较

比较项	Figshare	Dryad	PANGAEA
学科领域	工程科学、生命科学、自然科学、人类和社会科学	医学	环境和地球科学、生命科学
年代	2011	2008	1993
国别	英国	美国	德国
数据内容类型	标准的办公文档、图像、表格、原始数据、自由文本、存档数据	标准的办公文档、科学和统计数据、自由文本、结构文本	标准的办公文档、图像、自由文本、存档数据、音视频数据
数量(截至 2013 年 4 月 15 日,数据来源于 Datacite)	27 301 个数据文件	8 723 个数据文件	674 962 个数据文件
数据质量管理	是	是	是
数据唯一标识符	DataCite DOI	DataCite DOI	DataCite DOI
数据提供者	是	是	是
数据接口	检索 API	OAI - PMH、检索 API	OAI - PMH
数据许可协议	CC - BY 和 CC0 许可协议	CC 和其他	CC(大部分 CC - BY)
数据空间限制	1GB 的私有空间,无限制的公共空间	每个文件 1GB,每个数据包 10GB	
期刊数据存档的指定仓储	PLOS 的 7 种期刊	采纳 JDAP 策略的期刊	Elsevier 的 2 种期刊
长期保存	与 CLOCKSS 合作	与 CLOCKSS 合作	PANGAEA 负责

(1) Figshare^[33]

Figshare 为科研人员提供发布研究产出的平台,以便研究成果可以更好地被引用、共享和发现。Figshare 接受图表、媒体(包括音频)、海报 Poster、论文(包括预印本)和多文件(文件集)、数据集等所有类型的文件,为所有的对象分配 DOI,采用 Creative Commons 许可协议共享数据。Figshare 采用 Amazon 基于云的研究数据管理系统,来保证数据存储的安全和可靠性。PLOS 出版社 7 种期刊的附加数据也将存储到 Figshare 平台中,使得用户在浏览 PLOS 文章内容的同时也可以查看文章中的数据。F1000 Research 和 Figshare 正在合作研制 Widget,为用户提供更友好的方式来预览、下载、引用和共享 F1000 Research 文章的补充数据集。

(2) Dryad^[27]

Dryad 国际数据仓储库由期刊出版社、科研团体和其他利益相关者共同管理,以确保数据仓储库的可持续性。采纳 JDAP 策略的期刊,经常推荐 Dryad 作为提交科研数据的仓储。在 Dryad 中,可以提交生物科学领域出版文章的附加数据文件、软件或对于文章很重要的其他文件,对数据格式没有严格的要求,提交的数据文件都拥有永久的可解析的 DOI 标识。Dryad 与期刊无缝合作,简化提交的流程,将文章和数据提交整合在一起,作者可以增加描述数据文件的关键词,以及其他的描述数据文件的说明文件,以便于数据可以被再利用。Dryad 记录了数据的下载和查看次数,还与 TreeBase、KNCB、NCBI GenBank、DataONE 结成合作伙伴,相互之间可以进行数据交换。

(3) PANGAEA^[34]

PANGAEA 作为开放获取数据仓储,提供免费的数据存储和访问服务,目的是为了保存、发布和分发来自地球研究系统的地理参考数据。在 PANGAEA 中提交的数据需要经过质量控制,以标准的机器可读的格式存档,以提高数据的可重用性和互操作性。每个数据集都可以通过数字对象标识符(DOI)被识别、共享、发布和引用。数据可以作为出版物的补充进行存档,或作为可引用的数据集。PANGAEA 是 Elsevier 的 Earth Science Journals 和 Earth System Science Data (ESSD) 期刊指定的数据存储系统,PANGAEA 与 Elsevier、AGU、Nature、Springer、Wiley 等出版社合作,实现科学文献与 PANGAEA 数据之间的连接。

从表 2 中可以看出,开放科学数据仓储一般都为数据分配唯一标识符(DOI),以保证在可信赖的环境中提供数据集的持续永久性访问。采用 CC 协议来共享数据,提供数据开放接口,这些都有利于科学数据的开放共享和再利用。同时,为了保证数据质量,都采取一些质量控制机制。积极与出版社或者其他数据仓储合作,促进文献与数据之间以及数据与数据之间的关联和互操作。

6 开放科学数据的发现服务

为了提高科学数据的访问,改善在互联网上获取科学数据的简易性,帮助用户发现、访问和再利用数据,一些数据库商、搜索引擎开始对科学数据进行索引,提供科学数据的访问和发现服务。

(1) 通用搜索引擎

一些通用的搜索引擎,如 Google、Microsoft Academic Search、Scopus、BASE 等开始从 DataCite、PANGAEA 等抓取数据的元数据,并将数据和其他资源类型一起提供检索服务。BASE 是最大的开发获取资源搜索引擎之一,由比勒费尔德大学(德国)图书馆研制。截至 2013 年 4 月 27 日,BASE 中共收录 1 004 259 个数据集^[35]。

(2) 数据引文发现服务

2012 年 6 月,汤森路透(Thomson Reuters)公布了发现全球数据集的数据引文索引(Data Citation Index)^[36],提供跨学科和跨区域的科学数据访问,并对数据仓储中的数据集与同行评议的期刊、图书和会议论文文献进行关联,还可以跟踪研究数据的使用情况。汤森路透对数据仓储进行评估和筛选后,获取大约 70 个数据仓储的元数据(如 DataCite),超过 200 万的数据研究和数据集。并与数据存储库进行合作(如 ICPSR),获取书目记录和引用参考,来促进数据的可见性和透明度,以及评价数据的影响力和重要性。

(3) 学科数据发现服务

Web 上有很多免费的科学数据,这些原始数据一般不受版权保护,因此,出现了一些面向特定学科领域对这些数据进行下载和聚合并以增值资源方式重新发布的服务,如 ChemSpider、CrystalEye 等。ChemSpider^[37]由英国皇家化学学会(RSC)研制,从 400 多个数据源集成化合物数据,提供超过 28 万的化合物结构、属性

和相关信息的免费访问。提供文本和化学结构检索,使研究人员能够发现最全面的化学数据。CrystalEye由剑桥大学研制,提供关于晶体结构方面的开放数据服务。CrystalEye^[38]自动从出版商 Web 网站上获取论文的附加数据(如 RSC、ACS),截至 2011 年已获取约 150 000 篇文章的 250 000 个补充数据集,通过搜索工具提供方便的浏览和搜索服务。

(4) 文献与科学数据之间的关联服务

根据最近欧盟的 PARSE. Insight 研究显示^[39],超过 84% 的科学家认为将科学数据与论文进行关联是有用的。目前,在文献与数据之间建立关联大多都处于起步阶段,一些数据库开始提供这方面的服务,如从 PubMed 的文章连接到 GenBank、Dryad 的数据,从 Elsevier、Nature、Springer、Wiley、AGU 等的论文连接 PANGAEA 中的数据集。DataCite 和 CrossRef 也正在探索合作的可能性,使出版商可以识别数据集和文章之间的关系^[40]。

7 结 语

近年来,科学数据的数量呈指数增长,和研究论文一样,科学数据也被认为是一种重要的科研产出,大多数研究人员、资助机构、图书馆和出版商都认同开放和重用科学数据的好处,未来几年将是开放科学数据理念和应用的快速发展时期。越来越多的出版商与机构团体合作,积极实现数据和文献之间的整合,以互动的方式来展现数据,使公共存档的数据和文献之间可链接和可引用,提高数据的可检索性、发现性、可解释性和重新使用。图书馆作为信息资源和信息服务的提供者,可以开展科学数据的培训服务、科学数据的管理和存储服务、开放科学数据的发现和利用服务,促进科研数据的开放共享和重用。

(1) 开展科学数据的培训服务。科研人员很重视科研成果的发表,目前有很多商业期刊(如 Nature、Science)和开放期刊(如 PLOS One、Biodiversity Journal 等)都强制要求或者建议在发表论文时,同时提交支撑论文的数据或其他补充文件。另外,还出现了一些致力于数据出版的期刊(如 Geoscience Data Journal、Journal of Earth Systems Science Data、GigaScience 等)。这些期刊对提交的数据格式以及对数据元数据的描述都有一定的要求,并指定一些常见的数据仓储。图书馆可

以对相关的期刊以及数据仓储的数据存档政策按学科进行系统分析和整理,为用户开展科学数据的发表和存档政策的培训。

(2) 开展科学数据的管理和存储服务。科研人员更愿意将数据存放在数据仓储中,而不是出版社。图书馆可以建立数据仓储,鼓励科研人员提交和存档研究数据,分步骤分阶段实现数据共享,可以借鉴哈佛大学的 Dataverse Network^[28]和康奈尔大学图书馆的 DataStaR (Data Staging Repository)^[41]的经验。首先帮助科研人员将数据存储在机构库中,提供数据管理工具,然后允许已经公开的数据及元数据被学科仓储库或网络搜索引擎收割保存,促进数据更大范围的共享。

(3) 开展开放科学数据的发现和利用服务,对开放科学数据进行组织、整合和连接,使科学数据更容易被研究人员发现。图书馆可以在为用户提供的发现服务中,增加对开放科学数据的访问。按照主题领域收集开放的科学数据,包括数据仓储、数据期刊,建立开放科学数据的学科导航,将科学数据元数据集集成到图书馆目录或网站中,并建立数据和文献之间的连接。

参考文献:

- [1] 张晓林. 开放获取、开放知识、开放创新推动开放知识服务——30 会聚与研究图书馆范式再转变[J]. 现代图书情报技术, 2013(2): 1-10. (Zhang Xiaolin. Open Access, Open Knowledge, and Open Innovation Pushes for Open Knowledge Services——30 Convergence and a New Paradigmatic Shift for Research Libraries [J]. *New Technology of Library and Information Service*, 2013 (2): 1-10.)
- [2] Paving the Way to an Open Scientific Information Space: OpenAIREplus - linking Peer-reviewed Literature [EB/OL]. [2013-03-20]. <http://www.openaire.eu/en/component/content/article/326-openaireplus-press-release>.
- [3] iCORDI [EB/OL]. [2013-03-20]. <https://www.icordi.eu/Pages/Home.aspx>.
- [4] LERU. Open Research Data [EB/OL]. [2013-03-20]. <http://www.leru.org/files/general/Open%20Access%20to%20Research%20Data-FINALdocx.pdf>.
- [5] Open Data [EB/OL]. [2013-03-20]. http://en.wikipedia.org/wiki/Open_data.
- [6] Open Knowledge Foundation. Is It Open Data? [EB/OL]. [2013-03-20]. <http://www.isitopendata.org/guide/>.
- [7] OSTP. Increasing Access to the Results of Federally Funded Scien-

- tific Research [EB/OL]. [2013-03-20]. http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf.
- [8] Murray - Rust P, Neylon C, Pollock R, et al. Panton Principles, Principles for Open Data in Science [EB/OL]. [2013-03-20]. <http://pantonprinciples.org/>.
- [9] OECD Principles and Guidelines for Access to Research Data from Public Funding [EB/OL]. [2013-03-20]. <http://www.oecd.org/sti/sci-tech/38500813.pdf>.
- [10] Dissemination and Sharing of Research Results [EB/OL]. [2013-03-20]. <http://www.nsf.gov/bfa/dias/policy/dmp.jsp>.
- [11] NIH Data Sharing Policy [EB/OL]. [2013-03-20]. http://grants.nih.gov/grants/policy/data_sharing/.
- [12] NASA Data & Information Policy [EB/OL]. [2013-03-20]. <http://science.nasa.gov/earth-science/earth-science-data/data-information-policy/>.
- [13] RCUK Common Principles on Data Policy [EB/OL]. [2013-03-20]. <http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx>.
- [14] GHG - Europe Data Policy [EB/OL]. [2013-03-20]. http://www.ghg-europe.eu/fileadmin/user_upload/ghg/1_About_GHG/7_GHG_Data_Policy/GHG-Europe_Data_Policy_final.pdf.
- [15] PLOS One [EB/OL]. [2013-03-20]. <http://www.plosone.org/static/policies.action>.
- [16] BMC Evolutionary Biology [EB/OL]. [2013-03-20]. <http://www.biomedcentral.com/bmcevolbiol/about>.
- [17] Nature [EB/OL]. [2013-03-20]. <http://www.nature.com/authors/policies/availability.html>.
- [18] Joint Data Archiving Policy (JDAP) [EB/OL]. [2013-03-20]. <http://datadryad.org/pages/jdap>.
- [19] The High Level Expert Group on Scientific Data. Riding the Wave - How Europe Can Gain from the Rising Tide of Scientific Data [EB/OL]. [2013-03-20]. <http://www.grdi2020.eu/Repository/FileScaricati/c2194260-3ddf-47bd-93e4-68f8912a3564.pdf>.
- [20] Piwowar H, Vision T J. Data Reuse and the Open Data Citation Advantage [EB/OL]. [2013-03-20]. <https://peerj.com/preprints/1/>.
- [21] Henneken E A, Accomazzi A. Linking to Data - Effect on Citation Rates in Astronomy [EB/OL]. [2013-03-20]. <http://arxiv.org/abs/1111.3618>.
- [22] Sears J. Data Sharing Effect on Article Citation Rate in Paleogeography [EB/OL]. [2013-03-20]. <http://www.komfor.net/blog/unbenanntemitteilung>.
- [23] DataCite [EB/OL]. [2013-03-20]. <http://search.datacite.org/>.
- [24] Jones C A, Bouton K A, Hey J M N, et al. Data Publication; Outputs of the CLADDIER Project [EB/OL]. [2013-03-20]. http://www.pv2007.dlr.de/Papers/Jones_CLADDIER.pdf.
- [25] SageCite [EB/OL]. [2013-03-20]. <http://www.ukoln.ac.uk/projects/sagecite/>.
- [26] DCC. How to Cite Datasets and Link to Publications [EB/OL]. [2013-03-20]. <http://www.dcc.ac.uk/resources/how-guides/cite-datasets>.
- [27] Dryad [EB/OL]. [2013-03-20]. <http://www.datadryad.org>.
- [28] Dataverse Network [EB/OL]. [2013-03-20]. <http://thedata.org/citation/standard>.
- [29] Gesis [EB/OL]. [2013-03-20]. <http://www.gesis.org/en/services/data-analysis/data-archive-service/citation-of-research-data/>.
- [30] OAD [EB/OL]. [2013-04-10]. http://oad.simmons.edu/oadwiki/Data_repositories.
- [31] re3data [EB/OL]. [2013-04-10]. <http://www.re3data.org>.
- [32] Databib [EB/OL]. [2013-04-10]. <http://databib.org>.
- [33] Figshare [EB/OL]. [2013-04-15]. <http://figshare.com>.
- [34] PANGAEA [EB/OL]. [2013-04-15]. <http://www.pangaea.de>.
- [35] BASE [EB/OL]. [2013-04-15]. <http://www.base-search.net>.
- [36] Data Citation Index [EB/OL]. [2013-03-10]. http://wokinfo.com/products_tools/multidisciplinary/dci/.
- [37] ChemSpider [EB/OL]. [2013-03-10]. <http://www.chemspider.com/>.
- [38] CrystalEye [EB/OL]. [2013-03-10]. <http://wmm.ch.cam.ac.uk/crystaleye/>.
- [39] PARSE. Insight: INSIGHT into Issues of Permanent Access to the Records of Science in Europe [EB/OL]. [2013-03-10]. http://www.parse-insight.eu/downloads/PARSE-Insight_D3-6_InsightReport.pdf.
- [40] CrossRef Joins STM - DataCite Statement [EB/OL]. [2013-03-20]. <http://www.crossref.org/01company/pr/news081012.html>.
- [41] Steinhart G. DataStaR: A Data Sharing and Publication Infrastructure to Support Research [J]. *Agricultural Information Worldwide*, 2011, 4(1): 16-20.
- (作者 E-mail: huangyw@mail.las.ac.cn)