

德国国家科技图书馆 科学内容 DOI 注册中心研究进展

Jan Brase

(德国国家科技图书馆, 德国汉诺威)

摘要: 本文介绍了科学数据注册的背景及“原始科学数据出版与引用”项目概况, 研究了科学数据唯一标识符定义和结构, 分析了德国科技图书馆中科学数据的注册过程及所需要的基础设施, 论述了科学数据的可引用性, 并给出了3种其他科学内容注册的实例, 总结了该中心科学数据 DOI 注册的现状和下一步发展的方向。

关键词: 数字对象唯一标识; 科学数据; 德国科学技术图书馆

中图分类号: G356.5 文献标识码: A DOI: 10.3772/j.issn.1674-1544.2008.01.007

The German National Library of Science and Technology as a DOI – Registration Agency for Scientific Content

Jan Brase

(The German National Library of Science and Technology, Hannover, Germany)

Abstract: The paper introduced the background of the research of scientific data registration and the general situation of the project “Publication and Citation of Scientific Primary Data”, made a research of the definition and structure of DOI for scientific data, analyzed the scientific data registration process and required infrastructure of TIB library catalogue, then it discussed the citability of scientific data and gave registration examples of three other scientific content, finally summarized the status quo and development direction of the DOI registration agency (TIB).

Keywords: Digital Object Identifier (DOI), scientific data, TIB (German National Library of Science and Technology)

从2005年开始, 德国国家科技图书馆 (TIB) 成为正式 DOI 注册机构。公益性研究产生的科学内容, 大部分是原始科学数据集, 由德国国家科技图书馆赋予其数字对象唯一标识符, 可以在全球网络浏览器中进行访问。

1 科学数据的注册

科学文献传播的知识是原始科学数据经过一系列加工处理后的结果。这些数据经过分析、

综合和解释,最终成为科技文献发表。只有很少一部分的原始数据会在一般性的科学期刊上发表。尽管目前已经制定了一些有关数据归档的政策,但是在实际中,数据仍主要存储在公开的档案中,并没有存储在安全的机构库中。无法充分获取科学数据已经成了开展学科间国际化研究的一大障碍,从而可能导致不必要的重复研究,且查证结果也非常困难(Dittert, Diepenbroek & Grobe, 2001年)。虽然每年都花费了大量的研究资金,但是现有数据仍无法得到充分利用(Arzberger, Schroeder, Beaulieu, Bowker, Casey, Laaksonen et al. 2004年)。为进一步促进数据引用,鼓励好的科学实践及对科学研究的认可,德国研究基金会(DFG)发起了“原始科学数据的出版和引用”项目(Publication and Citation of Scientific Primary Data)。

德国国家科技图书馆是国际数字对象唯一标识基金会(IDF)的成员,业已被确定为科学原始数据数字对象唯一标识(DOI)的注册机构。该馆将从地球科学领域着手,开展数字对象唯一标识的注册工作。有关地球科学、气候和环境研究的原始数据都本地化存储在负责数据评估和维护的机构中。除提供本地数据外,德国国家科技图书馆还保存了可获取这些数据的网址,包括所有书目的元数据。在数据注册后,德国国家科技图书馆将为数据提供一个数字对象唯一标识,作为数字环境中内容对象的一个唯一标识。Lautenschlager 等对科学基础设施的内容做了详细的阐述^[1]。

2 数字对象唯一标识

2005年5月,德国国家科技图书馆正式成为数字对象唯一标识的注册机构。他们为所有数据都赋予了数字对象唯一标识。数字对象唯一标识就是为数字网络中使用的所有实体数据指定名称。它们可以用来提供当前的信息,包括它们(或有关它们的信息)在互联网上的位置。数字对象的信息可能会随着时间的推移而发生变化,包括查找的位置,但是它的数字对象唯一标识是永恒不变的。

数字对象唯一标识由两部分组成:前缀和后

缀。例如科学数据的数字对象唯一标识:10.1594/WGCC/IPCC_EH4_OPYC_SRES_B2_MM。其中,10.1594是前缀,确定这个数字对象唯一标识属于在德国国家科技图书馆注册的某个科学数据集;WGCC代表其研究机构(本例为世界气候数据中心);紧随其后的是世界气候数据中心内部的数据记录名称。数字对象唯一标识使用了美国国家研究推进机构(CNRI)的句柄系统(Handle System),可在全球所有的网络浏览器中解析。国际数字对象唯一标识基金会的网页上就嵌入了句柄服务。例如,网址http://dx.doi.org/10.1594/WGCC/IPCC_EH4_OPYC_SRES_B2_MM就可以解析这个数字对象唯一标识。另外,还可以在互联网浏览器上安装免费插件实现解析,从而方便地在浏览器地址栏中输入数字对象唯一标识。

3 图书馆目录中的科学数据

所有数据集在德国国家科技图书馆注册后,元数据就被储存起来。内容元数据是以定义明确的科学数据应用规范为基础的。应用规范包括按照针对电子媒体引用的ISO 690-2标准以及都柏林核心标准元数据属性确定的所有元数据。Brase 对所使用的元数据做了详细分析^[2]。

所有注册的数据集都可以通过相关数据中心的目录获取。一些出版物附带的辅助数据集,也被收录到德国国家科技图书馆在线图书馆目录里(图1)。

4 基础设施

设立专门的基础设施,可以灵活地为数据集注册数字对象唯一标识,并把元信息移植到相关的图书馆目录中。其关键要素就是网络服务。这也是德国国家科技图书馆提供自动和人工上载注册信息的中间件的一部分。网络协议和数据格式均基于XML,便于开发商理解。网络服务在传输层上使用了HTTP/HTTPS,可以通过多种常见的防火墙进行安全防护,而无需更改防火墙的过滤规则。

由于德国国家科技图书馆的网络服务采用

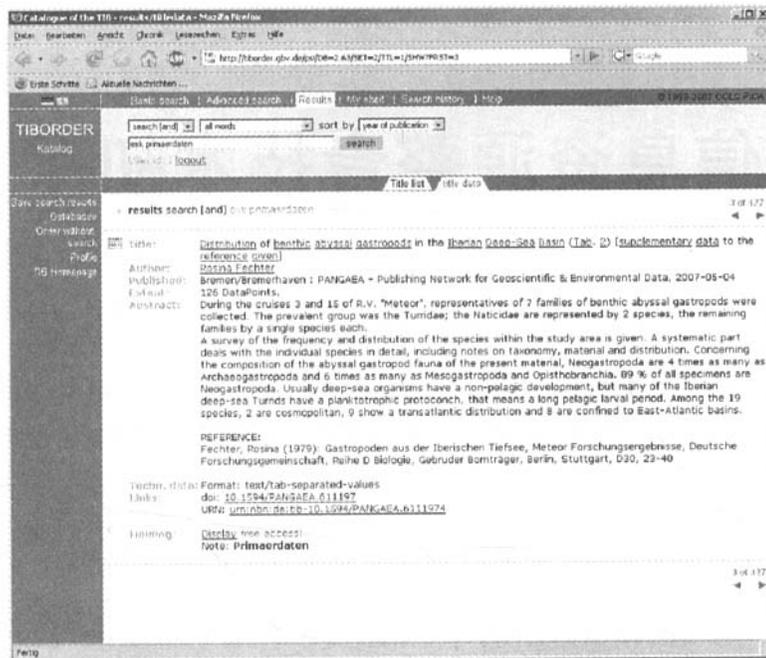


图1 在线图书馆目录中的数据搜索集搜索结果

了 SOAP(简单对象访问协议),因此,数据提供者可以把 WSDL(网络服务描述语言)文件输入到应用服务器中,进而把客户端代码嵌入中间件。

Hinzmann 对该基础设施进行了描述^[3]。

5 可引用性

由于所有数据都采用了数字对象唯一标识,致使原始科学数据不仅是科学出版物的一部分,而且还有了自己的身份。数字对象唯一标识就是这个原始数据自己的唯一身份。科学家在研究中可以通过数字对象唯一标识引用数据。如果一位科学家看到了使用被注册数据的文章,那么他可以对数据进行多方面的分析,并在自己发表的文章中通过该数据的数字对象唯一标识来引用这个数据。

由于学术研究能否得到认可通常是通过统计科学家的研究论文被引用次数进行考量的,因此收集数据也是作为学术研究的一个重要组成部分来完成的。一些文章的作者已经开始在参考书目中通过数字对象唯一标识引用数据。例如, Lorenz 等在文章使用和引用的数据就包括: Stendel, M. T. Smith, E. Roeckner, U. Cubasch (2004): OECHAM4_OPYC_SRES_A2: 110 years coupled A2 run 6H values, WDCC, doi:

10.1594/WDCC/EH4_OPYC_SRES_A2。上面所提到的是使用数字对象唯一标识来引用数据集的典型例子^[4]。

6 其他类型的内容

2006年,德国国家科技图书馆把注册的范围延伸至其他学科领域,比如包括医学、化学以及其他像晶体结构和灰色文献等类型的内容。例如:

(1) 病例研究。与欧洲变态反应与临床免疫学会(EAACI)合作,例如: M. Blanken, L. Bont (2007)。一名呼吸短促和困难的8个月大的女婴的数字对象唯一标识为 doi: 10.1594/eaacinet2007/CR/6-210907。

(2) 灰色文献。与欧洲计算机图形学会(Eurographics)合作,例如: K. Debattista, L. Paulo Santos, A. Chalmers, 并联组件渲染加快发光缓冲,欧洲计算机图形学会并行图形和可视化研讨会(2006年)第27-34页的数字对象唯一标识为 doi: 10.2312/EGPGV/EGPGV06/027-034。

(3) 晶体结构。与英国图书馆联网办公室(UKOLN)合作: 6, 7, 9, 10, 12, 13, 15, 16 - Octahydro-benzo-1, 4, 7, 10, 13-pentaoxacyclopent-

(下转第64页)

站之间实现互相链接;实现现有大型科学仪器中心在网站层次上的互联互通。

(3) 从深度和广度上加强大型仪器信息资源建设。大仪中心信息资源建设是大仪中心信息共享平台的基础,在深度建设方面,主要是完善内容,对仪器性能、原理、使用等进行完善而有深度的介绍;在广度方面,增加科普、论坛、机时预约、科研成果和相关文献介绍等内容。

(4) 推广离子探针中心的服务模式。在提供传统仪器信息服务的同时,开展远程共享和仪器使用情况监控等服务,不仅利用国内的仪器资源,还通过远程共享技术利用国外仪器资源为国际服务,从而提升大仪中心信息共享平台的服务功能;该模式具有极强的推广价值,并有力地提高了国内仪器服务在国际上的影响力。

(5) 在现有基础上建立国家大型仪器中心英文网站,提供仪器资源的英文信息,使大型仪器资源在世界范围内实现资源共享。该平台不仅能聚合现有的信息资源,而且能整合国内外与大型科学仪器相关的信息资源。国家投入巨资建设的大仪中心,其服务范围应该不仅仅是在中国,还

应该走向世界。从某种意义上讲,大仪中心代表了我国科学仪器使用的最高水平。

(6) 加强大仪中心信息共享机制的研究,通过大型科学仪器中心共享网络,发布仪器信息和相关合作研究信息,形成一个信息交流和同行合作的窗口,并通过合理的信息共享机制来吸引更多科学家、研究人员来大型科学仪器中心进行科研工作,产生高水平的科研成果^[5]。

参考文献

- [1] 董诚,张渝英,等. 国家大型科学仪器中心建设的思考[J]. 实验技术与管理, 2007(10): 4-5.
- [2] 《国家大型科学仪器中心建设发展规划研究》报告[R]. 中国科学技术信息研究所.
- [3] 王静. 我国科学仪器共享迈出关键步伐[N]. 科学时报, 2007. <http://www.cas.cn/html/Dir/2007/11/23/15/40/00.htm>.
- [4] 大型仪器中心共享与合作网[EB/OL]. <http://www.ccc-science.cn>.
- [5] 冉文清. 大型科学仪器协作共用平台建设与探索[J]. 现代仪器, 2006(4): 58-59.

(上接第 39 页)

tadecin(一种分子式结构 $C_{14}H_{20}O_5$) 的数字对象唯一标识为 doi: 10.1594/ecrystals.chem.soton.ac.uk/145。

7 现状

截止到 2007 年 10 月,德国国家科技图书馆已经注册了 47.5 万个数据集、1.25 万个科幻电影片段、6302 个病例研究、342 份技术报告以及 112 个学习对象。今后,对社团资助的研究所产生的科学内容进行登记将是德国国家科技图书馆的一项主要任务。这将包括对各种不同类型内容的注册,像晶体结构、地球取样以及 3D 模型等。

参考文献

- [1] Lautenschlager, M. Diepenbroek, M. Grobe, H. Klump, J. and Paliouras, E. World Data Center Cluster "Earth System Research" - An Approach for a Common Data Infrastructure in Geosciences. EOS, Transactions, American

Geophysical Union, 86(52, Fall Meeting Suppl.): Abstract IN43C-02, 2005.

- [2] Brase, J. 'Using digital library techniques - Registration of scientific primary data'. (LNCS 3232) Research and advanced technology for digital libraries, 2004.
- [3] Hinzmann, J. Erweiterung des XML Web Development Frameworks Cocoon zu einem Webservice am Beispiel der Registrierung von Primärdaten. B.Sc. Thesis, Universität Hannover, Hannover, Germany, 2005:82.
- [4] Lorenz, S.J. Kasang, D. Lohmann, G. 'Globaler Wasserkreislauf und Klimänderungen — eine Wechselbeziehung' Warnsignal Klima: Genug Wasser für alle?, Wissenschaftliche Auswertungen, Hamburg, Germany: 153-158.

参考数据

- [1] Stendel, M. T. Smith, E. Roeckner, U. Cubasch (2004): 0ECHAM4_OPYC_SRES_A2: 110 years coupled A2 run 6H values, WDCC, doi: 10.1594/WDCC/EH4_OPYC_SRES_A2.