

我国科学数据共享网站评价研究

赵伟¹ 彭洁^{1,2} 杨行¹

(1. 中国科学技术信息研究所, 北京 100038; 2. 武汉大学信息管理学院, 湖北武汉 430072)

摘要: 针对我国166个科学数据资源开放共享网站, 包括国家和地方科技基础条件平台科学数据平台、中科院科学数据库系统以及野外观测台站网站、自然科技资源网络共享平台等, 基于可见性、可得性、可用性三维评价体系, 进行了分指标、分类型、分领域和综合的绩效评价, 以期由政府等决策主体和资源建设机构提供信息支持。

关键词: 科学数据; 共享; 网站; 绩效评价

中图分类号: G203

文献标识码: A

DOI: 10.3772/j.issn.1674-1544.2014.02.001

Evaluation on Sharing Websites of Chinese Scientific Data

Zhao Wei¹, Peng Jie^{1,2}, Yang Xing¹

(1. Institute of Scientific and Technical Information of China, Beijing 100038; 2. School of Information Management, Wuhan University, Wuhan 430072)

Abstract: As one of the strategically resources of promoting national innovation, the network sharing platforms of scientific data have active function to make scientific data efficiently shared and utilized. In view of the 166 open-access sharing websites of scientific data in China, including the national and local scientific data sharing platforms, the database systems of Chinese Academy of Sciences (CAS), the website of the Field Observation Station of Ecosystems, as well as the sharing platform of natural scientific and technological resources, a comprehensive performance evaluation was carried out. The performance evaluation for sharing platforms of different types from different areas was completed based on a three-dimensional evaluation system, Visibility-Availability-Usability, in order to provide information support for decision-making bodies.

Keywords: scientific data, sharing, website, performance evaluation

1 引言

科学数据资源是国家科技创新的重要战略性资源之一。国际上, 科学数据管理和应用的发展趋势是在共同遵守的数据共享原则下建立全球性和区域性的数据共享网络, 系统采集、整理和存储数据, 使数据源不断增加和完善。近十年来我国开展了不同层面上的科学数据共享网站的建设

与运行^[1], 为推动我国科学数据共享共用、提高资源利用效率发挥了积极作用。

但是, 对这些科学数据共享网站现状的研究还十分有限, 尤其缺乏对网站数据资源、信息构建以及强调面向用户的服务水平等进行深入研究和综合考察。基于网站这一公开界面提取信息并进行系列评价, 一方面将有助于揭示目前国内科学数据资源开放共享工作的整体状况并为政府决

作者简介: 赵伟* (1975—), 女, 副研究员, 研究方向: 科技资源管理与共享。彭洁 (1965—), 女, 研究员, 研究方向: 科技资源管理、信息资源管理。杨行 (1990—), 女, 硕士研究生, 研究方向: 信息资源管理。

收稿日期: 2014年2月22日。

策提供有效的信息支持；另一方面系统化的评价结果也可以为用户使用网站资源及服务提供指导和帮助，推动公共服务，同时积极引导科学数据共享网站的建设和有效运行。因此，有必要对目前尚在运行的科学数据共享网站进行公开、客观的评价。

2 评价指标体系

目前，学者们专门针对科学数据资源共享网站评价开展了相关研究和实践^[2-4]。姚一鸣通过问卷调查的方法，选择科学数据共享功能指标和网站建设质量指标两个一级指标对地球系统科学数据共享网进行评估^[5]；樊秀娥等从系统质量、信息质量、服务质量、使用、用户满意度和个人影响等6个角度对科学数据共享系统的应用情况进行评估^[6]；李海燕等从资源建设质量、共享标准符合性、平台建设质量以及项目管理水平等4个维度构建了医药卫生科学数据共享工程质量评估的指标体系^[7]；戴琼洁从基础设施、人员配备、科学数据价值、服务能力和综合绩效等方面对陕西省科学数据共享平台进行了绩效评价^[8]。

此外，各类科学数据共享平台也出台了相应的评价制度和评价体系，如《地震科学数据共享项目评价制度细则》要求从资源建设、系统建设、运行管理、共享绩效与影响、发展前景等5个方面对地震科学数据共享中心及分中心的数据共享项目进行评价^[9]，而《教育部科技基础资源数据平台评估规则》中的评价指标主要包括资源建设、共享贡献、运行管理、发展前景等4个方面，两者大致相同^[10]。国家人口与健康科学数据共享网也开展过自评估。其评估分为数据中心评估指标体系和数据集评估指标体系两部分，评估指标主要包括资源建设、标准建设和服务建设等。

基于已有研究成果，本文拟从科学数据共享网站评价涉及的本源问题出发，构建一套科学可行的评价指标体系。彭洁等从“资源共享”的本质和内涵（以满足用户需求为判断准则）入手，从个体共享资源的行为过程出发，

并结合用户体验理论，提出了科技资源开放共享评价的“可见性－可得性－可用性（VAU）”模型（Visibility－Availability－Usability模型，简称“VAU模型”）^[11]。赵伟等认为该模型与用户获取资源、资源建设和管理环节都形成了良好的映射关系，并分别针对科技信息资源和科技实物资源的属性特征，提出了细化的一般性评价指标和特征性评价指标^[12]。基于此，本研究进一步结合科学数据以及科学数据共享网站的特征，初步构建了科学数据共享网站评价的二维和三维指标，并采用人工神经网络（BP）方法对指标进行优化，得到科学数据共享网站评价的指标体系（表1）。

其中，可见性主要从资源公开和资源可见程度两个方面进行考察，用来反映科学数据共享网站在推动科学数据资源的属性信息、状态信息，甚至是资源本身被用户方便看到的程度；资源公开维度主要考察数据资源的规模、公开率以及公开范围；资源可见程度主要以网站外部链接来衡量。可得性是指获取或使用科技数据资源及服务的难易程度，即共享的科学数据资源和服务能够有效地被得到、被使用；数据资源的可得性往往由多个要素支撑，这里主要从元数据质量高低、信息构建优劣和网站性能指标等3个方面进行评估。可用性的评价主要考察用户通过科学数据共享网站获取资源或服务后所能够指导自己工作或研究的便利程度；基于当前科学数据共享网站的发展程度及可用信息，这里主要从辅助信息及文档的丰富度、网站功能可用性以及数据更新情况等3个方面进行度量。

3 评价对象

目前，共享领域的科学数据资源主要是具有开放属性的自然资源类数据，包括资源环境科学、地球科学、生物、气象、农业、医药卫生、林业、海洋等；其他一些领域公开共享的数据资源较少，这与其领域内的信息化程度不高及信息公开程度有限等有较大关系。从网站类型来看，此次纳入评价对象的科学数据共享网站主要有5个渠道：54个国家科技基础条件平台科学数据平

表1 科学数据共享网站评价指标体系

一级指标	二级指标	三级指标
可见性	资源公开	资源规模
		数据公开率
		资源公开范围
	影响力	外链数
可得性	元数据质量	元数据完整度
		核心元数据填充率
	信息构建	导航便捷性
		搜索便捷性
	网站性能	首屏时间
		页面下载速度
		网站稳定性
可用性	数据文档与政策法规	数据文档丰富度
		共享政策及法规文档丰富度
	网站功能可用性	常用功能
		特色功能
	数据更新情况	数据更新率

台、46个中科院科学数据库系统、46个生态系统野外观测台站网站、13个自然科技资源网络共享平台以及7个省一级资助或内部资助构建的科学数据共享平台。本研究将对我国目前运行基本正常的166个科学数据共享网站的情况进行考察和评价。需要指出的是，这里的“科学数据共享网站”是广义的数据网络平台。本研究认为对于野外台站以及自然科技资源而言，虽然从资源类型上看，它们属于科技实物资源，并非传统观念里的科学数据共享平台或网站，但通过对实物资源采取信息化手段而形成的网络平台上，实际揭示和利用的是实物资源的描述性数据以及促进资源共享的相关信息。因此，本研究将该类资源共享网站也纳入研究范围。

全部评价对象中采用独立域名的网站数量不到有域名网站数量的1/5，可见国内大多数科学数据共享网站都是依托于某一工程项目或大的资源体系构建而成；非标准端口和网址路径的出现说明部分网站仍处于试运行状态或者仅是上级网站的子栏目。数据资源建设机构的地理位置分布极不均衡，41.6%的网站集中在北京地区，广东、湖北、山东、江苏、湖南等地也搭建了一定数量和规模的共享网站。

从承建单位来看，在166个评价对象中有81个网站的承建单位是中科院下属的研究所。可见，中科院是我国科学数据开放共享工作的主力军。中国农业科学院农业信息研究所、中国科学院寒区旱区环境与工程研究所、中国科学院地理科学与资源研究所和中国科学院新疆生态与地理研究所等单位承建了4个以上的科学数据共享网站。研究院所积极投身于共享网站的建设工作，但是同一单位不同网站数据资源的原创性及重复性情况还有待进一步考察。

4 总体评价

参与评价的166个科学数据共享网站中得分最高的是地球系统科学数据共享网（82.54分）。能达到60分以上水平的网站只有51个，占全部网站的30.72%，所有网站的平均得分为54.17分；只有42.8%的网站高于平均水平。从总体上看，国内科学数据共享网站的整体水平还有待进一步提高。

4.1 不同类型网站间的对比

为进一步考察国内科学数据共享网站间的差异和距离，本文对国家科技基础条件平台科学数据平台、中科院科学数据库系统、自然科技资源

类网络共享平台、生态系统野外观测台站网站以及地方科学数据共享平台的平均得分进行了分类统计,分析结果见图1。

从一级指标来看,自然资源类共享平台的可见性最好,生态系统野外观测台站的可得性最好,而可用性表现最好的是国家科技基础条件平台科学数据平台;从不同类型网站最终得分来看,总分在60分以上的绝大多数网站来自国家科技基础条件平台科学数据平台或中科院科学数据库系统;生态系统野外观测台站和地方科学数据共享平台的总分整体偏低。

4.2 不同学科领域间的对比

不同学科领域科学数据共享网站的得分也存在较大差异。就总分而言,地球科学、气象、医药卫生领域表现要优于资源环境、林业以及海洋科学领域。除化学学科相关共享网站可见性指标

的平均得分要比可得性指标的平均得分高之外,其余学科共享网站的可得性指标表现均优于可见性指标。可得性平均得分较高的是农业科学和气象科学,而可用性表现最好的则是地球科学领域,资源环境领域相关网站的可用性得分整体偏低(图2)。

5 分项指标评价

5.1 可见性

可见性总体排名前三位的共享网站分别为国家人口与健康科学数据共享平台、国家农业科学数据共享平台和国家材料环境腐蚀科学数据共享服务网。

(1)资源公开方面。①资源规模。共享网站公开的资源量按照数据库和数据集两种组织方式分别统计,以数据集为组织方式的平台数为84

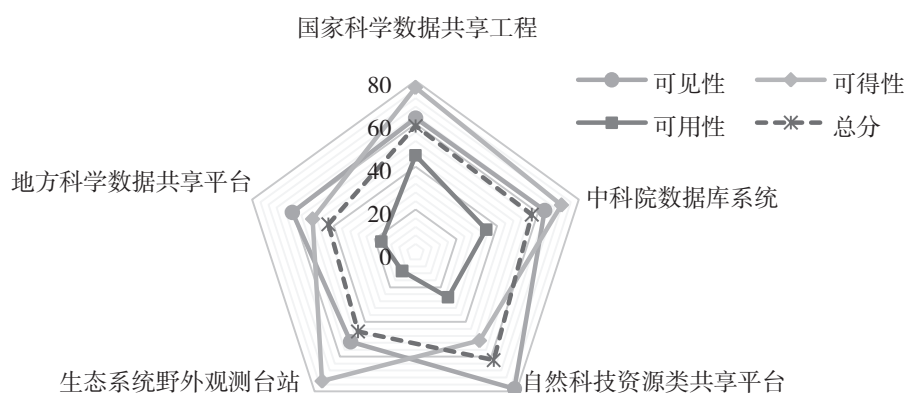


图1 不同类型共享网站一级指标的得分情况

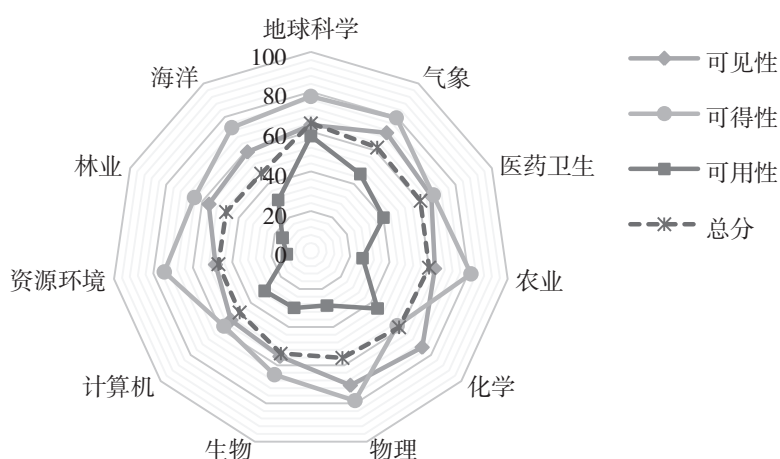


图2 不同学科领域科学数据共享网站综合得分及一级指标对比

个,其资源量主要集中在(0,200)个的区间中;以数据库为组织方式的平台数为82个,其资源量主要集中在(0,100)个的区间中。在166个共享网站中,当前可以统计到的数据量(包括数据集和数据库)为24568个,平均一个网站拥有148个数据集(库);公开数据量在1000以上的平台只有4个,分别是:CERN中心、CERN综合中心、交通科学数据共享网和地球系统科学数据共享网。②数据公开率。全体评价对象数据公开率的平均值为83.79%,其中有79个网站的数据公开率为100%;从学科领域来看,化学、物理、气象以及农业领域数据公开率的平均值都在95%以上。③资源公开范围。绝大多数网站都需要用户注册登录后才提供数据检索、下载或评论功能。

(2)影响力方面,本文主要采用外链数指标来考察。被评价网站的外部链接数目整体上呈现两极分化的现象,近1/3的网站都没有外部链接,只有25个网站的外链数高于平均值。

从不同类型网站在可见性三级指标上的平均得分来看,自然资源类共享平台在可见性各三级指标上的表现都比较突出,中科院科学数据库系统在资源公开范围上表现较好,生态系统野外观测台站的资源规模普遍较小,地方科学数据共享平台数据公开率的平均值总体偏低。

5.2 可得性

可得性排名前三位的网站分别为渔业和水产业科学数据中心、空间科学数据共享平台、林业平台云南省分中心。

(1)元数据质量。元数据有利于数据的交换和用户对数据资源的识别,在科学数据共享网站中具有重要作用。在全部评价对象中,36个网站没有专门编制元数据(其中31个平台的资源组织方式为数据库),部分网站的元数据质量较差,这对科学数据共享网站的标准化和规划化将会产生很大影响。

(2)信息构建。一些网站放置了许多与科学数据共享无关的内容,其主页的有效信息较少;还有一些网站缺乏高效的导航功能和检索服务,用户很容易在网站中“迷路”,无法快速找到目

标数据资源。

(3)网站性能。一些资源量较大、数据质量较好的网站由于没有考虑在其他网络条件下平台的访问速度和网站稳定性,导致用户无法快速甚至正常使用网站资源和服务,从而在很大程度上会影响用户对其服务的满意程度和对网站的粘着度。

从不同类型网站在可得性二级指标上的平均得分来看,国家科技基础条件平台科学数据平台的元数据质量最好,自然资源类共享平台在性能指标上的表现较好;不同类型共享网站在信息构建指标上的表现相差不大,生态系统野外观测台站在该指标上的平均得分最高。

5.3 可用性

科学数据共享网站的可用性情况整体上并不理想,全体评价对象可用性指标的平均得分为29.83,能达到60分线的平台只有17个,仅占全部评价对象的10.24%。可用性排名前三位的网站分别为:地球系统科学数据共享网、极地科学数据共享平台和气象科学数据共享网。

(1)数据文档与政策法规。大部分网站对其提供的数据资源都有单独的数据描述,只有16个网站除元数据描述外,还提供专门的数据描述文档或使用说明,可以进行下载查看,帮助用户更好的利用数据资源。在以数据集为资源组织方式的网站中,以地球系统科学数据共享网为总平台的地球科学类平台做得比较好。以数据库为资源组织方式的共享网站中,中国科学院化合物参考数据库、国际科学数据服务平台以及基础医学科学数据中心和公共卫生科学数据中心表现较好。在共享政策法规方面,地震科学数据共享中心、药学科学数据中心以及林业科学数据共享中心提供相关信息服务较多,既包括宏观政策及数据共享机制、措施或管理办法,也包括数据共享相关标准规范。

(2)网站功能可用性。常见问题解答和咨询服务是科学数据共享网站中最常见的两个服务功能,极地科学数据共享平台、气象科学数据共享网等平台还提供在线数据分析或可视化服务等特

色功能。

(3)数据更新情况。数据更新指标在一定程度上能够反映建设单位对网站的重视程度和维护情况,在166个评价对象中多达114个平台没有任何关于网站数据更新的相关信息或栏目。

从不同类型网站在可用性二级指标上的平均得分来看,国家科技基础条件平台科学数据平台在辅助文档及信息方面的工作比较突出,自然资源类共享平台在数据更新方面的表现较好,网站功能指标平均得分最高的则是中科院科学数据库系统。相对于其他类型的网站,生态系统野外观测台站和地方科学数据共享平台在可用性方面的表现都不太理想,它们在网站数据更新方面的工作尤其需要重视和加强。

从总体上看,有些共享网站在可见性、可得性和可用性3个方面均表现较好,如地球系统科学数据共享网、气象科学数据共享网和极地区域数据共享运行服务中心,其平台和资源建设比较成熟。但也有的网站在某个或者多个方面做得不足,甚至网站处于停滞的状态,网站的建设和运行还处于初级阶段,与最初的预期目标还有一定差距。

6 网站管理建议

根据本文对我国166个科学数据共享网站开展的评价,对科学数据共享网站的管理提出如下建议:(1)在科学数据共享网站建设方面:①需要尤其注重提升元数据质量以及搜索引擎功能的可见性;②对于首屏时间、页面下载速度以及网站稳定性等作为网站绩效考核的重要指标,需要在技术支持方面给予保障。(2)在网站机制建设和组织管理方面:①建议加强网站的统一管理和协调工作,建立和强化相应的协调机制和激励机制;②加强数据的分类规范及其实施,将共享深入到各资源类型实际业务当中去,同时也吸引更多的机构加入到科学数据共享工作中来;③形成常年性基于第三方的评价机制,激励科学数据共享网站建设逐渐完善;④建立优质科学数据资源

索引目录,通过一定的评审程序,对特定数据中心的科学数据资源进行专业性评价,质量优秀的可以将元数据发布到专业的科学数据索引目录中,对其进行唯一标识的注册并实现增值利用;⑤加强科学数据的国际化,逐步建立科学数据引用机制,推进科学数据在更大范围内获得共享和有效利用。

参考文献

- [1] 刘润达.科学数据门户及其构建实践研究——以地球系统科学数据共享网为例[D].北京:中国科学院,2009.
- [2] Miranda Gonz Lez F J, Ba Egil Palacios T M. Quantitative Evaluation of Commercial Web Sites: An Empirical Study of Spanish Firms [J]. International Journal of Information Management, 2004,24(4):313-328.
- [3] 陆宝益.网络信息资源的评价[J].情报学报,2002,21(1):71-76.
- [4] Benbunan-Fich R. Using Protocol Analysis to Evaluate the Usability of a Commercial Web Site [J]. Information & Management, 2001,39(2):151-163.
- [5] 姚一鸣.科学数据共享网站评估体系研究——以地球系统科学数据共享网为例[D].北京:中国科学院研究生院工程教育学院,2008.
- [6] 樊秀娥,龚琼宇.科学数据共享系统评估方法的探讨[J].中国卫生工程学,2009(3):179-180.
- [7] 李海燕,崔蒙.医药卫生科学数据共享工程质量评估模式与指标体系框架研究[J].世界科学技术-中医药现代化,2009,11(4):578-581.
- [8] 戴琼洁.陕西省科学数据共享平台运行机制研究[D].西安:西安电子科技大学,2011.
- [9] 中国地震局.地震科学数据共享项目评价制度细则[S/OL]. [2008-05-20]. <http://www.escience.gov.cn/Show-Article2.jsp?id=3079&r>.
- [10] 教育部.教育部科技基础资源数据平台评估规则(试行)[S/OL].[2012-04-26]. <http://www.dost.moe.edu.cn/dostmoe/jgsz/zcfg/20120426101556>.
- [11] 彭洁,赵伟,屈宝强.基于用户视角的科技资源开放共享评价理论模型研究[J].中国科技资源导刊,2013(2):2-6.
- [12] 赵伟,彭洁,屈宝强,等.网络环境下科技资源开放共享评价指标体系的构建[J].中国科技资源导刊,2013(3):10-15.