

基于re3data的英国科学数据发布平台研究^{*}

张莎莎, 黄国彬, 耿睿

(北京师范大学政府管理学院, 北京 100875)

摘要: 本文以re3data为数据获取源, 选取英国247个科学数据发布平台为研究对象, 从责任主体、平台功能、数据资源、数据传输四个角度出发, 对英国科学数据发布平台特点进行研究, 总结英国科学数据发布平台特点及建设经验, 为科研人员访问和利用现有科学数据发布平台、科学数据发布平台建设者建设和完善平台提供一定指导。

关键词: 科学数据; 发布平台; 英国; re3data

中图分类号: G250

DOI: 10.3772/j.issn.1673-2286.2017.06.003

1 引言

随着“互联网+”时代的来临, 科研人员每天在科学研究中产生海量科学数据。科学数据, 又被称作科研数据或研究数据, 指在科研活动中产出的能够反映客观世界本质、特征及变化规律的原始数据, 以及根据科学研究活动需要而加工处理的数据集合^[1]。当前, 国内外科学研究步伐日益加快, 多渠道收集与使用科学数据已成为十分重要的研究方法, 然而由于科学数据来源广、数量大、种类多等特点, 科学数据的获取和处理已成为科研人员面临的重大难题之一。而解决该难题的关键则是建设一个集数据存储、描述、共享、获取等功能于一身的科学数据发布平台, 至此, 不同类型、各具特色、质量参差不齐的科学数据发布平台应运而生。

根据德国研究基金会资助构建的re3data (一个汇集全球不同学科科学数据发布平台的注册表) 统计, 截至2017年4月6日, 共有1 839个科学数据发布平台在此注册^[2], 其中美国 (894个)、德国 (281个)、英国 (247个) 的科学数据发布平台数量排名前三。对众多科学数据发布平台的选择和使用, 成为科研人员面临的难题之一。鉴于统计工作量及语言限制等因素, 本文选取英国的247个科学数据发布平台为例, 对其特点进行总结和分析, 辅助科研人员选择适合的科学数据发布平台,

以减少重复研究、提高科研效率、增加科研收益, 并对我国研究机构在组织建设和完善自身科学数据发布平台方面提供指导和借鉴。

2 科学数据发布平台国内外研究现状

本文以“科研数据”OR“科学数据”OR“研究数据”为关键词, 并以“平台”OR“仓储”OR“共享”等为关键词, 在中国知网、Web of Science等数据库进行文献检索, 通过对现有文献的分析来揭示有关科学数据发布平台的研究现状。

2.1 国内研究现状

目前, 国内有关科学数据发布平台的研究主要集中于三方面。(1) 介绍国外各领域具有代表性的科学数据发布平台案例。覃丹选取英美两国一流高校构建的5个社会科学数据管理与共享平台 (ROPER、SSDA、ODUM、ICPSR、UK Data Archive), 从整体结构、检索服务、导航服务、咨询/帮助服务、教育培训服务等方面介绍各平台建设现状^[3]; 陈秀娟等调研化学领域科研数据发布平台的创建国家、建立时间、平台类别、开放情况等, 选取Cambridge Structural Database、

^{*} 本研究得到国家社会科学基金项目“云计算环境下图书馆信息资源安全政策法律研究” (编号: 11CTQ004) 资助。

ChemSpider和ChemSynthesis来深度分析化学数据发布平台的服务内容^[4]; 李翼等考察国外四大医学科学数据发布平台(PMC、Europe PMC、AIIHW、CT)的建设现状、共享情况和元数据特点, 并将其与国内的人口与健康科学数据平台进行对比分析^[5]。(2) 研究科学数据发布平台的支撑软件。张计龙等从系统功能、二次开发可能性和系统可维护性等角度分析比较主流平台软件Dspace、Fedora Commons、Dataverse、Nesstar和在线分析软件SDA的特点^[6], 殷沈琴等对以上5款软件进行部署测试, 对系统的进阶功能、元数据标准、在线分析功能进行评估和系统选型, 提出构建高校科学数据管理平台的建议^[7]。(3) 探讨科学数据发布平台的设计与实现。于海军等针对载人航天空间应用与科学数据在线分发共享需求, 提出一种多层次、松耦合、易扩展的平台架构, 解决了平台数据逻辑模型设计、元数据自动提取、数据便捷在线服务、轨道仿真等技术问题及其实现方式^[8]; 周宝平介绍太原市科学数据共享平台的设计过程和开发技术, 指出该平台具有元数据汇交、数据发布、元数据检索及多种方式数据共享等功能, 具有较好的扩展性和复用性^[9]。

2.2 国外研究现状

通过对国外文献调研发现, 目前有关科学数据发布平台的研究较少且主要围绕两个主题。(1) 研究现有的科学数据发布平台。Chourasia等从功能和技术维度探讨科学数据发布平台SeedMe的特点, 其中, 功能维度包括数据文件和集合属性、程序文件发布工具、元数据、下载、访问控制、安全获取等内容, 技术维度包括内容管理、安全控制、Web服务、视频编码等内容^[10]; Kraft等从建设模式、建设目标、系统架构、元数据标准等方面介绍Radar, 其主要面向无数据共享传统学科的科学数据发布平台^[11]; Herrick等调研了可公开访问的医学成像数据存储库XNAT Central, 并从当前服务、托管数据、贡献数据、未来发展等方面对其深入介绍^[12]。(2) 研究科学数据发布平台的构建或改进方法。Herzinger等开发SmartR插件, 该插件有利于增强科学数据发布平台tranSMART的互动和动态视觉分析能力^[13]; Huang等利用和扩展Chaste和sif软件框架开发新科学数据发布平台, 该平台有助于促进研究者间科学数据共享, 有利于共享科学数据访问^[14]; Kruger等提出归档和检索实时科学数据的一般架构(即科学

数据发布平台), 并从元数据角度探索改进科学数据发布平台存储和检索功能的方案^[15]。

目前有关科学数据发布平台的研究存在两点不足:

(1) 国内外学者都倾向于研究个别实践案例, 没有从宏观角度出发, 对现有科学数据发布平台的特点进行梳理; (2) 现有研究虽从元数据标准、数据服务、软件平台等角度对科学数据发布平台进行介绍, 但欠缺从平台语言、数据类型、学科分布等角度的探索。因此, 本文选取英国247个科学数据发布平台为例, 从责任主体、平台功能、数据资源、数据传输四个角度出发, 梳理英国科学数据平台的整体特点。

3 英国科学数据发布平台的特点分析

语种的多样性对信息的传播和获取具有促进作用, 就英国的247个科学数据发布平台而言, 其使用了43种语言且每个平台都使用英语这一语种; 此外, 232个平台(占94%)的语言类型仅为英语, 15个平台(占6%)采用两种及以上语言, 其中Data Hub^[16]语言版本最多(见表1)。英国科学数据发布平台管理机构在采用英语作为平台主要语言的同时, 逐渐采用多语种版本来提高平台可读性, 该举措虽目前普及度不高, 但对于母语非英语的研究者大有裨益。除平台语言通用、多语种外, 英国的科学数据发布平台在其责任主体、平台功能、数据资源、数据传输方面皆具有显著的特点。

3.1 责任主体

3.1.1 来源国家/组织

来源国家/组织指科学数据发布平台的责任主体所属国家或组织。通过对英国247个科学数据发布平台调研发现^[2], 有107个科学数据发布平台(占43%)仅由英国建设或管理, 其余平台由英国和其他国家/组织合作建设或管理(见表2)。在合作建设或管理的科学数据发布平台中, Consortium of European Social Science Data Archives的责任主体所属国家/组织最多(22个); 由2—10个国家/组织合作建设或管理的平台数目分别为71、38、13、4、5、2、4、1、1和1。从合作对象看, 合作国家/组织数量在1—3个的平台最多(占87%); 此外, 英国与美国、德国、法国、加拿大、澳大利亚合作最多, 这些国家正好也是科学数据发布平台数量排行前10的国家。

表 1 英国科学数据发布平台的语言类型

语言种类数/种	平台数量/个	具体语言类型	案例
1	232	英语	VecNet
2	4	英语/法语	ShareGeo Open
	4	英语/威尔士语、英语/丹麦语、英语/德语、英语/日语	DNA Data Bank of Japan
3	2	英语/威尔士语/法语、英语/西班牙语/葡萄牙语	Reading Experience Database
4	2	英语/德语/俄语/希腊语、英语/西班牙语/日语/中文	ComBase
5	1	英语/中文/法语/日语/约鲁巴语	International HapMap Project
13	1	英语/德语/法语/西班牙语/意大利语/马扎尔语/波兰语/葡萄牙语/俄语/波斯语/希伯来语/乌克兰语/中文	Galaxy Zoo
41	1	英语/葡萄牙语/日语/意大利语/捷克语/加泰罗尼亚语/西班牙语/法语/希腊语/塞尔维亚语/挪威语/斯洛文尼亚语/芬兰语/俄语/德语/波兰语/荷兰语/保加利亚语/韩语/马扎尔语/拉脱维亚语/阿尔巴尼亚语/中文/波斯语/塔加路语/蒙古语/土耳其语/越南语/乌克兰语/希伯来语/克罗地亚语/泰语/印地语/丹麦语/印尼语/瑞典语/冰岛语/高棉语/罗马尼亚语/阿拉伯语/立陶宛语等	Data Hub

注：“/”用以分隔同一科学数据发布平台使用的不同语言，“、”用以分隔不同的科学数据发布平台使用的语言。

3.1.2 管理机构

管理机构是负责科学数据发布平台运营的直接单位。调研发现^[2]，44个平台（占18%）属于一家机构独立管理模式，其余203个平台（占82%）由多家机构联合管理，通常合作机构数为2—5家（占59%）。英国的科学数据发布平台注重合作建设模式，以吸引合作伙伴加盟，在资源、资金、技术、服务等方面各显其长，共同提供优质数据服务。如UK Data Archive由埃塞克斯大学成立并提供技术和部分数据资源支持，英国的经济与社会研究委员会、国家档案馆、联合信息系统委员会等提供资金支持等^[17]。此外，从平台管理机构性质看，4个平台（占2%）仅由盈利性机构管理，如GeoNames；15个平台（占6%）由盈利性和非盈利性机构共同管理，如Crystallography Open Database；其余228个平台（占92%）由非盈利性机构管理，如UK Data Service。

3.2 平台功能

3.2.1 开发软件

开发软件是支撑科学数据发布平台运作的基础架构和硬件系统。除154个平台的开发软件信息未知以外，剩余93个平台的开发软件信息如图1所示。

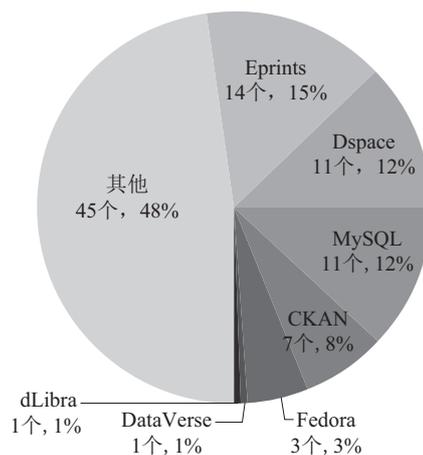


图 1 英国科学数据发布平台的开发软件情况

英国科学数据发布平台通常基于EPrints、Dspace、MySQL、CKAN、Fedora、DataVerse、dLibra软件开发。上述软件基本属于开源软件，具备科学数据发布平台所需的存储、发布、检索、下载等数据管理功能，可免费获取且支持二次开发或系统移植；此外，它们获取便利、使用简易、安全可靠、兼容性强，满足数据管理所需的上传、下载、检索、筛选、分享、社交等功能。如DataVerse具有在线统计与分析功能，CKAN提供浏览、下载次数统计和5星级评价标注等功能。从开发软件角度看，平台的功能大部分取决于开发软件的选择。

表 2 英国科学数据发布平台的合作国家/组织分布

个

国家/组织数量	平台总数量	具体国家/组织	平台总数量	案例
1	107	英国	107	National Geoscience Data Centre
2	71	英国/美国	27	SureChEMBL
		英国/欧盟	16	Health Improvement Network
		英国/国际组织	8	University of Salford Data Repository
		英国/比利时	6	EUROCAT
		英国/德国	4	NMRshiftDB2
		英国/加拿大	3	GOBASE
		英国/奥地利、英国/丹麦、英国/日本、英国/卢森堡、英国/西班牙、英国/澳大利亚、英国/瑞士	7	OPEN-ADAS
3	38	英国/美国/欧盟	10	Ensembl Bacteria
		英国/美国/瑞士	7	Universal Protein Resource
		英国/美国/国际组织	5	HUGO Gene Nomenclature Committee
		英国/德国/欧盟	3	DEPOD
		英国/美国/德国	3	ROSAT Mission
3	38	英国/欧盟/国际组织	2	European Nucleotide Archive
		英国/美国/意大利、英国/美国/日本、英国/美国/爱尔兰、英国/美国/澳大利亚、英国/美国/加拿大、英国/瑞士/欧盟、英国/瑞士/国际组织、英国/比利时/国际组织	8	Worldwide Protein Data Bank
4	13	英国/美国/欧盟/国际组织	3	FlyReactome
4	13	英国/美国/德国/欧盟、英国/美国/澳大利亚/瑞士、英国/美国/法国/欧盟、英国/美国/意大利/国际组织、英国/德国/法国/瑞典、英国/德国/俄罗斯/希腊、英国/德国/挪威/欧盟、英国/荷兰/欧盟/国际组织、英国/澳大利亚/南非/国际组织、英国/美国/意大利/卢森堡	10	ArrayExpress
5	4	英国/德国/瑞士/丹麦/欧盟	2	STRING
		英国/美国/加拿大/瑞典/加纳、英国/美国/德国/加拿大/立陶宛	2	INDEPTH Data Repository
6	5	英国/美国/澳大利亚/墨西哥/日本/希腊、英国/挪威/瑞士/法国/芬兰/比利时、英国/美国/加拿大/法国/南非/国际组织、英国/美国/日本/爱尔兰/荷兰/朝鲜、英国/奥地利/瑞士/芬兰/挪威/俄罗斯	5	ComBase
7	2	英国/德国/荷兰/法国/瑞士/瑞典/欧盟、英国/德国/丹麦/塞内加尔/贝宁/布基纳法索/欧盟	2	West African Vegetation
8	4	英国/美国/加拿大/法国/日本/比利时/德国/国际组织	2	Intermagnet
		英国/美国/德国/斯洛伐克/希腊/奥地利/捷克/瑞士、英国/法国/希腊/俄罗斯/意大利/德国/荷兰/欧盟	2	mzCloud
9	1	英国/美国/法国/新加坡/意大利/印度/加拿大/瑞士/欧盟	1	IMEx
10	1	英国/美国/德国/意大利/加拿大/法国/日本/澳大利亚/中国台湾/国际组织	1	International Mouse Phenotyping Consortium
22	1	英国/奥地利/捷克/爱沙尼亚/法国/希腊/德国/匈牙利/爱尔兰/立陶宛/西班牙/瑞典/瑞士/荷兰/挪威/丹麦/罗马尼亚/芬兰/意大利/卢森堡/斯洛文尼亚/欧盟	1	Consortium of European Social Science Data Archives

注：“/”表示同一科学数据发布平台由不同国家/组织建设或管理，“、”用以分隔不同的科学数据发布平台的来源国家/组织。

3.2.2 平台类别

根据平台的功能定位,可将科学数据发布平台分为学科知识库和机构知识库。除11个平台(占4%)未注明平台类别外,有170个平台(占69%)类型为学科知识库,如UniProtKB/Swiss-Prot主要存储和发布遗传学学科方面的科学数据;有35个平台(占14%)类型为机构知识库,如DataSTORRE主要存储和发布英国斯特林大学研究者的科学数据;有31个平台(占13%)属于混合型知识库,如UK Data Archive不仅存储和发布机构成员的科学数据,还存储和发布注册会员、人文社科领域的科学数据^[17]。

3.3 数据资源

3.3.1 学科分布

从整体看,平台学科分布涉及人类与社会科学、生命科学、自然科学、工程科学四大类,14个一级子类,45个二级子类(见表3)。从各类学科平台数量分布可见,生命科学和自然科学是英国科学数据发布平台较重视的两大类学科。本文将内容只涉及一种学科大类、一种一级子类且一种二级子类的科学数据发布平台称为单学科平台,其余为多学科平台。经过统计发现,除MRC National Survey of Health and Development未注明学科外,183个平台(占74%)为多学科平台,如National Geoscience Data Centre存储和发布的内容涉及水研究、大气科学与海洋学、地球物理学与测地学3个二级子类学科;63个平台(占26%)为单学科平台,如STRING中的内容涉及的学科为基本的生物和医学研究。

3.3.2 内容类型

英国科学数据发布平台发布的内容类型多种多样,至少14种(见表4)。从内容类型数量看,13个平台(占5%)发布的内容只有一种类型,如Chemical Database Service只发布科学和统计数据格式的内容;217个平台(占88%)发布的内容类型数量在2—7种,如EBI Metagenomics发布科学和统计数据格式、纯文本、图片、原始数据和结构化图形5种类型的内容;Loughborough Data Repository发布的类型内容最多,达14种。

3.3.3 元数据标准

有76个平台对所使用的元数据标准情况进行了说明,主要涉及13种元数据标准(见表5)。其中,*Dublin Core*是使用最多的一项标准,其次分别为*DataCite Metadata Schema*、*DDI*、*ISO 19115*。*Dublin Core*是描述信息资源的通用标准,其内容简洁明了^[18];*DataCite Metadata Schema*可应用于任意领域,准确且有效地识别出资源,以便于用户引用和检索^[19];*DDI*是社会科学领域数据标准,其设计不仅考虑获取信息,还考虑采用机器可读的格式去表达数据处理、发现及分析全过程^[20];*ISO 19115*是面向地理科学数据国际元数据标准,可为数字化地理科学数据集描述提供清晰的过程^[20]。有的科学数据发布平台不止采用一种元数据标准,如Environmental Information Data Centre使用3种元数据标准,分别为*Dublin Core*、*DataCite Metadata Schema*和*ISO 19115*。

3.4 数据传输

3.4.1 数据上传

大多数平台对数据上传设有限制条件,以确保数据质量。173个科学数据发布平台(占70%)允许有限的上传数据,64个平台(占30%)不面向公众上传数据,8个平台(占3%)面向用户开放上传数据。这3种方式并不互斥,可组合使用。如Virtual Liver Network的上传方式为不面向其他公众上传和有限制上传,只有机构成员才可以注册并在平台上存储发布数据^[21]。有限制上传包括注册上传、机构成员注册上传、注册付费上传。此外,96个科学数据发布平台(占39%)规定上传数据需遵循相应数据上传许可协议,如EMEP规定了两个上传许可协议,分别为Data Policy和Terms of Use。据此可知,为提高数据质量,英国多数科学数据平台采取有限制上传和制定上传许可协议等举措。

3.4.2 数据获取

大多数平台都在积极响应开放获取运动,以推动科学信息的广泛传播和无障碍地获取。207个平台(占84%)获取方式为开放,109个平台(占44%)获取方式为有限制获取,33个平台(占13%)获取方式为暂时禁

表 3 英国科学数据发布平台的学科分布

学科大类名称	一级子类名称	二级子类名称
生命科学 (177)	生物学 (128)	基本的生物与医学研究 (94)
		动物学 (25)
		植物科学 (20)
	医学 (86)	医学 (49)
		微生物学、病毒学、免疫学 (43)
		神经科学 (7)
农业、林业、园艺与兽医 (16)	农业、林业、园艺与兽医 (10)	
自然科学 (107)	地学 (包括地理学) (52)	大气科学与海洋学 (23)
		地理学 (14)
		地球物理学与测地学 (13)
	地学 (包括地理学) (52)	水研究 (10)
		地球化学、矿物学与结晶学 (9)
		地质学与古生物学 (5)
	化学 (34)	分子化学 (8)
		生物化学与食品化学 (6)
		方法开发 (化学)、分析化学 (5)
		物理与理论化学 (3)
		化学固体与表面研究 (2)
		聚合物研究 (1)
	物理 (20)	天体物理学与天文学 (9)
		原子、分子与等离子体的光学、量子光学与物理 (7)
		粒子、核与场 (5)
		凝聚态物理 (3)
数学 (6)	数学 (1)	
人类与社会科学 (67)	社会与行为科学 (30)	社会科学 (17)
		经济学 (10)
		教育科学 (4)
		心理学 (4)
		法学 (3)
	人文科学 (22)	历史 (11)
		古代文化 (5)
		美术、音乐、戏剧与媒体研究 (5)
		语言学 (5)
		文学研究 (4)
		神学 (2)
		非欧洲语言与文化、社会与文化人类学、犹太研究与宗教研究 (1)
工程科学 (39)	计算机科学/电气与系统工程 (9)	计算机科学 (7)
		电气工程 (2)
		系统工程 (1)
	热能工程/过程工程 (3)	工艺工程、技术化学 (2)
		热能技术、热机、流体力学 (1)
	材料科学与工程 (2)	材料科学 (2)
		材料工程 (1)
	建筑工程与建筑 (2)	建筑工程与建筑 (1)
	机械与工业工程 (1)	力学与建筑机械工程 (1)

注: 括弧内数字代表各分类数量。

表4 英国科学数据发布平台的内容类型

平台的内容类型	数量/个	占比
标准办公文档	150	61%
科学和统计数据格式	148	60%
纯文本	131	53%
图片	123	50%
原始数据	119	48%
结构化图形	115	47%
结构化文本	85	34%
归档数据	69	28%
应用软件	47	19%
视听数据	43	17%
数据库	36	15%
基于网络的数据	23	9%
源代码	10	4%
配置数据	3	1%
其他	103	42%

表5 英国科学数据发布平台的元数据标准使用情况

数据类型	数量/个	案例
<i>Dublin Core</i>	26	data.bris Research Data Repository
<i>DataCite Metadata Schema</i>	8	Cambridge Structural Database
<i>DDI</i>	7	Consortium of European Social Science Data Archives
ISO 19115	7	Environmental Change Network
<i>RDF Data Cube Vocabulary</i>	5	UniProtKB
<i>CF Metadata Conventions</i>	3	British Atmospheric Data Centre
<i>FGDC/CSDGM</i>	3	GeoCommons Archive
<i>ABCD</i>	1	The University of Oxford Text Archive
<i>CIF</i>	1	British Atmospheric Data Centre
<i>DIF</i>	1	British Oceanographic Data Centre
<i>ISA-Tab</i>	1	FAIRDOMHub
<i>OAI-ORE</i>	1	Imperial College High Performance Computing Service Data Repository
<i>Repository-Developed Metadata Schemas</i>	1	Archaeology Data Service
其他	11	Intermagnet

止获取,仅23个平台(占9%)获取方式为无法获取。同数据上传类似,不同数据获取方式也并不互斥。如Environmental Information Data Centre的获取方式为暂时禁止获取和注册获取,即所有数据需用户注册才有获取权限,另有部分数据需等到被解禁后才允许用户下载^[22]。

3.4.3 数据引用

数据引用指科学数据发布平台对用户引用平台上数据的相关要求、说明或限制条件。147个平台(占60%)制定有引用指南,如Dryad要求科学数据施引者在引用时注明数据来源和数据来源文献,并对引用格式进行范例说明^[23];100个平台(占40%)暂未制定引用指南,如STRING。由此可见,英国科学数据发布平台较重视保护科研人员的学术成果。对科学数据予以引用不仅是对科学数据贡献者的尊重和对成果的认可,还有助于学者追根溯源,了解更多有关被引用的科学数据内容或验证相关研究的科学性和有效性,进而推动科学技术的不断发展和进步。

4 结语

通过以上研究发现,英国科学数据发布平台除了语言类型以英语为主,多语言为辅外,还具有3方面的特点。(1)合作建设或管理。英国科学数据发布平台大多是国家间、机构间合作建设或管理的模式,这种模式可以博采众长,充分发挥各组织机构特点,为研究者提供优化数据服务。(2)平台功能受开发软件和平台类别影响。平台开发软件除具有上传、下载、检索、筛选、分享、社交等功能外,不同的开发软件还有其特色功能,如在线统计、分析、评价等;不同类别的平台(学科知识库或机构知识库),其功能定位不同,目前英国科学数据发布平台主要倾向于发布某一个或多个学科文献资源。(3)数据资源内容广、类型多、有标准。英国科学数据发布平台在学科上基本涵盖所有主流和特色学科,数据资源丰富齐全;在内容类型上,其格式多样且以机器可读的类型为主,方便研究者应用和处理;在元数据标准上,其涉及的元数据标准多样,以国际通用为主,提高数据交换效率。(4)数据上传有要求、数据引用有规范、数据获取较容易。为对平台科学数据质量加以控制,多数平台对数据上传操作具有限制要求,如注

册上传、机构会员上传或付费上传等。英国大多数科学数据发布平台都在积极践行开放获取运动,对公众免费开放。在数据引用方面,英国目前有超过半数的科学数据发布平台都制定了引用指南,有效规范引用程序。

科学数据发布平台是支撑科研活动的关键基础设施,对建设经验丰富的英国科学数据发布平台特点进行研究,不仅有助于我国科研机构吸收先进经验,建设适应我国科研实情的科学数据发布平台,同时对于国内研究者选择适合的平台存储、共享或利用科学数据也具有积极作用。

参考文献

- [1] 司莉,邢文明.国外科学数据管理与共享政策调查及对我国的启示[J].情报资料工作,2013,34(1):61-66.
- [2] re3data.org[EB/OL].[2017-03-20].<http://www.re3data.org/search>.
- [3] 覃丹.英美社会科学数据管理与共享服务平台调查分析[J].图书情报工作,2014(16):67-75,142.
- [4] 陈秀娟,吴鸣.学科领域科研数据知识库调研与分析——以化学领域为例[J].图书情报工作,2015(9):111-118,147.
- [5] 李翼,吴丹.开放医学科学数据平台调查研究[J].图书情报工作,2015(18):24-29,50.
- [6] 张计龙,殷沈琴,张用,等.社会科学数据的共享与服务——以复旦大学社会科学数据共享平台为例[J].大学图书馆学报,2015(1):74-79.
- [7] 殷沈琴,张计龙,张莹,等.社会科学数据管理服务平台系统选型研究——以复旦大学社会科学数据平台为例[J].图书情报工作,2013(19):92-96.
- [8] 于海军,李盛阳,焦鹏,等.一种空间应用与科学数据分发共享服务平台的设计与实现[J].载人航天,2016(4):494-499.
- [9] 周宝平.论科学数据共享平台的设计与实现[J].山西科技,2010(2):48-49.
- [10] CHOURASIA A,WONG M,MISHIN D,et al.SeedMe:a scientific data sharing and collaboration platform[C]//Xsede16 Conference on Diversity,Big Data,and Science at Scale.[S.1.]:ACM,2016:48.
- [11] KRAFT A,RAZUM M,POTTHOFF J,et al.The radarproject:a service for research data archival and publication[J].ISPRS International Journal of Geo-Information,2016,5(3):28.
- [12] HERRICK R,HORTON W,OLSEN T,et al.XNAT central:open sourcing imaging research data[J].Neuroimage,2016,124(Pt B):1093.
- [13] HERZINGER S,GU W,SATAGOPAM V,et al.SmartR:an open-source platform for interactive visual analytics for translational research data[J/OL].Bioninformatics,2017.[2017-04-03].<https://www.ncbi.nlm.nih.gov/pubmed/28334291>.Doi:10.1093/bioinformatics/btx137.
- [14] HUANG J,ZHANG X,EISENHAEUER G,et al.Scibox:online sharing of scientific data via the cloud[C]//2014 IEEE 28th International Parallel and Distributed Processing Symposium, Phoenix, AZ,2014:145-154.
- [15] KRUGER A,LAWRENCE R.An architecture for real-time warehousing of scientific data[C]//International Conference on Scientific Computing,Csc 2005,June,Las Vegas,Nevada.[S.1.]:DBLP,2005:151-156.
- [16] datahub.Open knowledge international[EB/OL].[2017-04-05].<https://datahub.io/en/>.
- [17] University of Essex.About us[EB/OL].[2017-04-01].<http://www.data-archive.ac.uk/about>.
- [18] 赵华,王健.国内外科学数据元数据标准及内容分析[J].情报探索,2015(2):21-24,30.
- [19] DataCite Schema.Metadata Schema4.0[EB/OL].[2017-04-05].<http://schema.datacite.org/>.
- [20] 刘峰,张晓林.科学数据元数据标准述评及其通用化设计研究[J].现代图书情报技术,2015(12):3-12.
- [21] The University of Manchester and HITS gGmbH.Browse by scale[EB/OL].[2017-04-06].<http://seek.virtual-liver.de/>.
- [22] FORD H,GARBUTT A,SKOV M.Coastal biodiversity and ecosystem service sustainability(CBESS)soil pH on salt marsh sites at morecambe bay and essex[EB/OL].(2016-05-03)[2017-04-06].<https://catalogue.ceh.ac.uk/documents/c726249f-c2d8-4aeb-9af2-60a40de40be2>.
- [23] North Carolina State University.Frequently asked questions[EB/OL].[2017-04-06].<http://datadryad.org/pages/faq#using>.

作者简介

张莎莎,女,1992年生,硕士研究生,研究方向:信息分析,E-mail: buhuishuide@126.com。

黄国彬,男,1979年生,博士,硕士生导师,副教授,研究方向:信息法学、信息分析,E-mail: huanggb@bnu.edu.cn。

耿骞,男,1965年生,博士,教授,研究方向:政府信息管理、知识组织,E-mail: gengqian@bnu.edu.cn。

Research on UK Scientific Data Publishing Platforms Based on Re3data

ZHANG ShaSha, HUANG GuoBin, GENG Qian
(School of Government, Beijing Normal University, Beijing 100875, China)

Abstract: This paper studies UK scientific data publishing platforms from four aspects: responsibility subjects, platform functions, data resources and data transmissions, and sums up the characteristics and construction experience of UK scientific data publishing platforms by taking re3data, a registry of global scientific data publishing platforms as the data acquisition source, and selecting 247 scientific data publishing platforms in the UK as the research object, to provide guidance for researchers' accessing and using the existing scientific data publishing platforms and builders' constructing and improving scientific data publishing platforms.

Keywords: Scientific Data; Publishing Platforms; UK; Re3data

(收稿日期: 2017-04-13)

2017年全国知识组织与知识链接学术交流会 征文通知

为了探讨网络环境下知识组织与知识链接的新理念、新思路、新方法,中国科学技术信息研究所、国家科技图书文献中心和中国科学技术情报学会拟于2017年9月在北京召开第八届“全国知识组织与知识链接学术交流会”。现在面向国内图书情报界及相关领域的专家、学者征文。具体会议时间、地点另行通知。

一、会议主题

1. 知识组织理论与方法研究
2. 知识组织系统构建与应用
3. 分类法、叙词表、本体等研究进展
4. 元数据管理及其标准应用
5. 大数据中数据的清洗、组织和分析
6. 数据挖掘、人工智能发展
7. 语义关联与语义出版
8. 数字图书馆资源发现系统与技术
9. 科学计量、引文分析与专利分析
10. 知识管理与知识服务
11. 云计算与微服务
12. 用户分析与个性化服务
13. 知识图谱及可视化分析
14. 其他

欢迎广大图书馆学、情报学、出版传播相关研究、教学人员,及相关信息组织技术人员踊跃投稿。会议将评选优秀论文,并由中国科学技术情报学会颁发优秀论文荣誉证书。优秀论文将推荐到《数字图书馆论坛》上发表。

二、征文要求

1. 文章要求:观点明确、主题突出;必须为未经发表的论文;统一用A4纸排版,以电子邮件方式提供Word格式文档;正文字数应控制在4 000—8 000字。
2. 来稿请提供:中英文题目、中英文作者及单位、中英文摘要和关键词、正文、参考文献;论文后请附作者简介,包括作者单位、联系电话、电子邮箱、通讯地址及邮政编码等。
3. 截稿日期:2017年7月31日。

三、联系方式

联系人:覃艳 吴雯娜

电话:010-58882369

征文邮箱:KOLink@istic.ac.cn

会议网址:<http://168.160.16.186/conference>