

# 开放引文语料库实践案例及启示\*

宋丹辉

(河南科技大学图书馆, 洛阳 471023)

摘要: 开放引文数据作为一种新的开放关联数据, 在国外已取得长足发展, 而在国内还未开始建设。为缩小差距, 本文以国外已发挥重要作用的开放引文项目为学习案例, 深入分析其在数据规模、技术路线、元数据模型和 workflow 方面的建设与实践, 详细阐述其在解决引文数据不易获取、不一致、语义缺乏等问题的方法与特征, 并从发展理念、关键步骤等方面总结经验, 以期为我国引文数据开放关联建设提供建议。

关键词: 引文数据; 开放关联; 开放引文语料库

中图分类号: G230

DOI: 10.3772/j.issn.1673-2286.2017.07.009

## 1 引言

引文数据是学术交流的重要元素, 也是引文分析的基础, 但来源受限, 大多存储在 Web of Science、Scopus 等少数几个商业引文数据库中, 较难免费获取。Web of Science 由美国科学家尤金·加菲尔德于 1964 年创建的引文索引发展而来, 其主要商业对手是 Scopus。二者都具有覆盖学科范围广、文献类型多样、引文数据丰富的特点。为访问其中资源, 用户需花费巨额经费, 还要遵守严格的保密协议。一些免费资源平台如 Google Scholar、Microsoft Academic Search、百度学术、必应学术等, 虽然支持文献引用统计、生成不同格式引文, 或通过软件导出等, 但都制定有严格的使用条款, 用户无法进行大规模引文数据采集, 从而无法自动构建引文网络, 开展计量、评价或智能化分析等研究<sup>[1-2]</sup>。

针对上述问题, 业界提出两种不同的解决方案: 一是引文数据开放存取。随着一系列研究报告及政府公文日发布, 推行开放存取政策已成为业界共识, 学者也将研究重点逐渐聚焦到实施效果上。实践表明, 将论文和数据以开放存取方式发布会提升其引用次数和学术影响力<sup>[3-6]</sup>。鉴于此, 有学者尝试引文数据的开放存取<sup>[7]</sup>, 以便于科研人员顺利追踪引用轨迹、深入开展科研工作, 并促进在线论文的发现、讨论和引用。二是把

引文数据发布为关联数据, 借助 URI 规范, 通过 HTTP/URI 机制链接到以 RDF/XML 编码的相关数据对象, 从而实现富链接和富语义效果, 这不仅有助于发现高被引论文、纠正错误引文数据, 还有助于开发新的智能应用。

随着开放存取运动及科学数据管理研究的不断发展和深入, 开放引文逐渐得到越来越多的关注, 开放引文数据库也逐渐受到科研人员的重视, 如开放引文项目 (Open Citations Project, OCP)<sup>[8]</sup> 和 Dryad 数据库<sup>[9]</sup>。前者由英国联合信息系统委员会资助, 于 2010 年启动, 旨在改变全球学术交流和学术出版的现状, 促进引文数据开放。该项目遵循 CC 协议, 并以 RDF 格式出版书目引用信息, 使得人们能够自由查看、研究、重用及丰富引文数据, 不受任何限制地遍历引文链接。后者由美国国家自然科学基金会资助, 于 2008 年 9 月启动, 旨在存放优质数据资源, 并促使科技文献中的数据发现和再利用。Dryad 数据库通过联合学术团体、出版社、研究机构、教育机构、基金资助机构等构建学术交流体系, 以促进学术文献中基础科研数据的保护和再利用。目前, Dryad 虽然已被许多主流期刊采纳, 但其特色主要体现在规范数据引用格式上, 相较于 OCP 的主要成果——开放引文语料库 (Open Citations Corpus, OCC), 其在数据关联方面还存在很大差距<sup>[10]</sup>。此外, 英国南安普

\* 本研究得到国家社会科学基金青年项目“基于引文内容标注的引文数据开放关联模型及发布流程研究” (编号: 17CTQ005) 资助。

顿大学、美国康奈尔大学及arXiv网站也于1999—2002年共同开展了“开放引文计划”，但由于各种原因目前已经无法使用。

截至目前，鲜有文献针对“开放引文项目”的技术路线、工作流程等进行研究。鉴于此，本文以OCP为切入点，讨论引文数据开放关联过程中面临的问题、可能的解决方案，并总结经验教训，以期为我国引文数据开放关联提供参考。

## 2 OCC的建设实践

引文数据是书目数据的子集，书目数据开放关联的成功经验可以为引文数据提供参考，基于开放获取和关联数据进行引文数据开放关联的探索逐渐提上日程，主要的开放引文项目代表是OCP，其主要成果为OCC<sup>[11]</sup>。OCC遵循CC协议，提供从学术文献中抽取的、用语义出版和引用本体（Semantic Publishing and Referencing Ontologies, SPAR）描述的、与OCC元数据模型一致的、准确的书目引用信息，用户可自由地对这些书目信息进行构建、拓展、丰富和重用，不受任何知识产权或数据库法律的限制。

### 2.1 OCC的规模

OCC于2011年发布第一版，实现对OA-PMC（PubMed Central Open Access）中204 637篇论文引文数据的开放关联，包括6 325 178条指向3 373 961篇论文的引文记录。该语料库的规模虽然不大，但基本涵盖1950—2010年生物医学文献索引中20%的内容，包括已在PubMed发表的生物医学领域的全部高被引论文。

随着OA-PMC馆藏数量的不断增加（截至2014年12月，馆藏资源数量已经是创建时的2倍），以及arXiv preprint server中引文数据的导入（在生物医学等基础上，增加了数学等学科的引文数据），OCC的覆盖范围也逐渐得到扩展，下一步将增加CiteSeerX、CitEc（Citations in Economics）等免费数据库的引文数据。截至2014年12月，CiteSeerX已积累近1 350万条引文记录，CitEc已积累1 354万余条引文记录。这3个数据库累计覆盖约198万篇论文的引文数据，但相对于科技文献总量而言，该数据量并不算大，OCC仍有很长的路要走。目前，OCC的工作人员正修改数据模型，加强基础设施建设，以进一步扩大覆盖范围。

### 2.2 OCC的技术路线

现有期刊的引文数据主要通过CrossRef收割，出版商将论文引文数据提交至CrossRef，支持免费引用关联服务。然而，CrossRef的使用条款是通过元数据描述的，且这些引文数据默认是不公开的，仅能在符合出版商使用条款的条件下通过元数据搜索服务获取。对于过刊而言，若想实现引文数据的开放获取，出版商必须专门向CrossRef发邮件说明。这种方法虽然简单直接，无成本，但整个进度受工作人员回复时间限制。总之，引文数据开放获取的操作效率太低，引入智能化自动处理技术才是根本解决之道<sup>[12]</sup>。

为解决上述问题，OCC设定5个逐次递进的任务：

- （1）创建语义基础设施，开发或重用能够满足学术创作和出版需求的、支持以RDF格式描述科技文献书目数据、引文数据的语义模型，如本体或RDFS（Resource Description Framework Schema）词汇表；
- （2）开发注释工具，允许作者基于上述语义模型用恰当的语义声明对文档进行语义增强；
- （3）扩展上述书目实体和书目引用的语义处理设施，以解决数据实体和数据引用问题；
- （4）通过具体实例示范本体应用于现实数据的具体过程，创建科研数据与书目实体的引用关系，及描述书目实体与科研数据间引用关系的RDF元数据；
- （5）将OA-PMC中所有论文的引文数据转化为RDF格式，并以开放关联数据的格式发布在OCC上，以便第三方组织免费开展创新性使用<sup>[13]</sup>。

### 2.3 OCC的元数据模型

为描述并关联OCC的各类书目实体，OCC构建专门的元数据模型，如图1所示。OCC元数据模型主要包括6类书目实体：bibliographic resources（fabio:Expression）、resource embodiments（fabio:Manifestation）、bibliographic entries（biro:BibliographicReference）、responsible agents（foaf:Agent）、agent roles（pro:RoleInTime）、identifiers（datacite:Identifier）。其中核心类为“fabio:Expression”，具备“title”“subtitle”“PublicationYear”“edition”等属性，描述其题目、出版年和版本等信息，其通过“frbr:part”与“biro:BibliographicReference”建立关联；“biro:BibliographicReference”又通过“biro:references”与“fabio:Expression”的关联，实现反馈，描述论文

与参考文献的关系；通过“datacite:hasIdentifier”与“datacite:Identifier”建立关联，描述论文及其引用数据的关系；通过“pro:isDocumentContextFor”与“pro:RoleInTime”建立关联，而“pro:RoleInTime”

通过“pro:isHeldBy”与“foaf:Agent”关联，描述论文是由代理机构采用某种创作方式产生；通过“fabr:endeavour”与“fabio:Manifestation”建立关联，描述内容表达与实体表现的关系<sup>[14]</sup>。

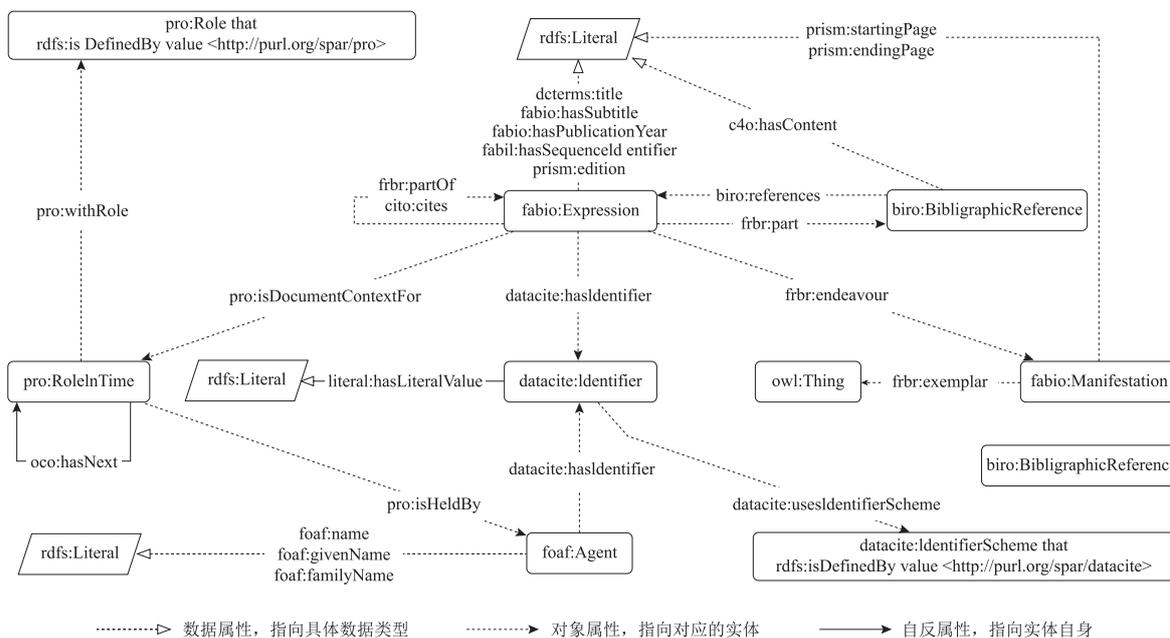


图1 OCC元数据模型中的主要本体实体<sup>[14]</sup>

OCC由若干个子数据集构成，每个子集对应一类书目实体，其中每个实体都被赋予唯一的URL（具体由“https://w3id.org/oc/corpus/”+“实体类名首字母缩写”+“/”+“子集内唯一序列码”构成），如“https://w3id.org/oc/corpus/be/537”代表“bibliographic entries”子集中第537个书目实体，可通过内容协商机制以HTML、RDF/XML、Turtle或JSON-LD等多种格式下载<sup>[15]</sup>。描述这些实体的来源信息元数据用PROV-O本体和PROV-DC扩展来表达。

为更好地对术语进行管理，相关术语都收集在开放引用本体（Open Citations Ontology, OCO）中，其是一个用于分组存放从若干现有本体中抽取出来的、互为补充的、书目实体元素的专属空间，以便为OCC提供描述性元数据。

## 2.4 OCC抓取实例数据的流程

如图2所示，在OCC中对具体引用数据的抓取工作主要由BEE（Bibliographic Entries Extractor）和SPACIN（SPAR Citation Indexer）实现，程序代码可从

OCC的GitHub软件存储库中下载。具体步骤为：（1）解析PMC论文的XML来源文档；（2）利用DOI和书目实体生成JSON文档；（3）对于每个引证或被引资源，若检测到相关ID（DOI、PMID或PMCID），则进一步确认该资源是否存在，若存在，则直接进行第5步；（4）若资源不存在，则从记录中抽取可能的IDs，并检索ORCID和CrossRef；（5）创建新的元数据资源，若CrossRef返回信息，则所有相关元数据都可以使用，否则只采纳基础元数据（IDs和记录）；（6）将所有声明加载至三元组存储库，并将其加入文档系统中以便将来恢复数据<sup>[16]</sup>。

BEE主要负责为每篇论文创建JSON文档。针对PMC中的每篇论文，BEE从可访问的XML来源文档中抽取该论文的所有元数据信息（包括所有可获取的唯一标识符，如DOI、ISSN、ISBN、ORCID、URL或Crossref member URL等）和参考文献信息（包括通过唯一标识符获取的信息），并将所有数据加入最终的JSON文档中。此外，JSON文档也包含数据出处、提供者、OCC维护者等信息。

SPACIN主要负责处理BEE生成的JSON文档，并

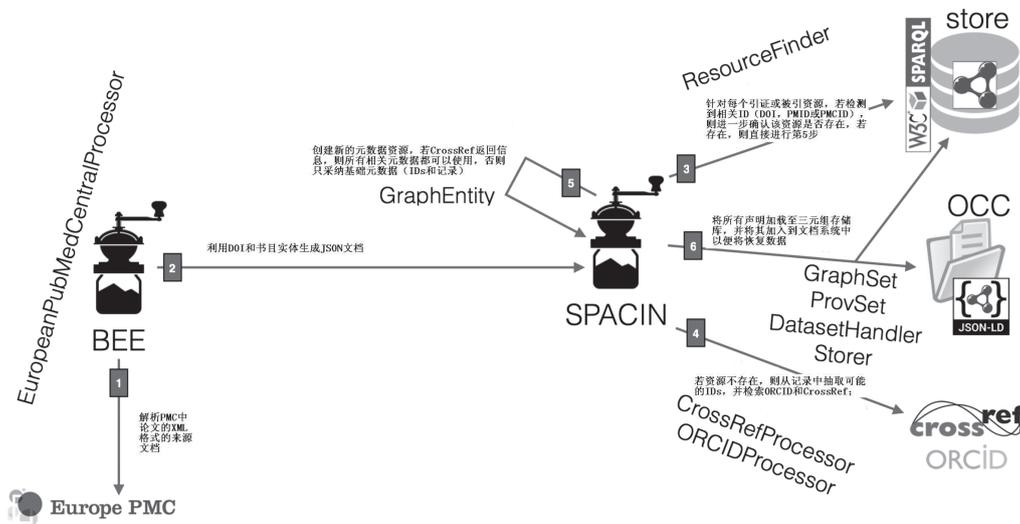


图 2 OCC 抓取实例数据的流程<sup>[16]</sup>

通过Crossref API和ORCID API检索附加的有关引证或被引文档的元数据信息。这些API也可以通过检索到的唯一标识符（如DOI、ISSN、ISBN、ORCID、URL或Crossref member URL等）来辨别书目资源和责任者。一旦检索到相关元数据，便可创建对应的RDF声明（若相关RDF声明已经存在，则重用便可），并以JSON-LD格式存储到文件系统和OCC三元组存储库中。此外，考虑到空间和性能因素，三元组存储库几乎包含维护实体的所有数据，除来源数据和数据集自身描述信息外，二者只能通过HTTP协议访问。

OCC抓取实例数据的整个流程是连续不断的，直到不再加载新的JSON文档为止。因此，OCC实例是随时间动态增长的，容易通过调整配置、与更多不同来源REST APIs交互来扩展PMC以外的论文，进而收集更多元数据和参考文献的信息。目前，OCC由博洛尼亚大学的计算机科学与工程学院来维护，用户可通过一个SPARQL查询端点和一个支持数据消费的浏览界面来访问。自2016年7月以来，该机构已经完成PMC中科技论文参考文献列表的抓取、处理与发布工作。自2016年9月，所有抓取到的数据均能以数据集方式下载。

### 3 OCC的实践特征

作为免费引文库，OCC打破现有科技文献引文数据的商业束缚，针对引文数据可获取性差、准确性不高、缺乏语义信息等问题提供一系列解决方案，改变了在数字化科研基础设施中的传统形象，成为引文数据开放关联的最佳实践。

#### 3.1 通过数据收割协议提高引文数据的可获取性

OCC的最终目标是汇集世界范围的科技文献以及艺术领域、人文学科领域的引文数据。但目前大多数出版商都把引文数据设为付费内容，受版权保护。在OCC的引领下，已有Nature等少数几家出版商把书目和引文数据开放共享。

针对现有引文库的商业垄断及半开放性控制，OCC与部分期刊出版商（如牛津大学出版社、麻省理工学院出版社等）就引文数据的日常收割问题签订合作协议，并计划进一步整合收割数据，以清晰展现论文、论文与数据库的引用关系，拓展作者、机构、基金、论文间的多维语义关系。OCC将允许用户免费浏览多种来源的引文数据，包括传统学术出版物及其他数据论文等，但所有引文数据都会标明来源<sup>[15]</sup>。

基于开放获取的引文数据，OCC计划开发多种智能化分析服务，如分面搜索和浏览工具、建议和趋势识别服务以及基于时间的可视化服务等，其中部分服务已在OCC原型系统中实现。随着覆盖范围的不断扩大，OCC在计算引用评价指标有用性方面也将不断提升。

#### 3.2 借助外部权威数据源纠正引文数据的不一致问题

受数据库自身收录文献类型、数量或范围的影响，不同引文库引文数据相差较大，计算机领域更为明显，同一篇论文在Google Scholar的引用次数会远高于其

他数据源。由于计算机领域很多重要成果以会议论文方式出版,而Google Scholar对会议论文的收录及标引最完整,因此,既使其准确性低于其他引文库,计算机领域学者也更倾向于使用Google Scholar查询论文的引用次数。但Google Scholar不允许非订购用户对其进行检索,并以定义的格式返回结果,这大幅降低了其在构建引文网络中的影响与地位<sup>[17]</sup>。

此外,引文数据在准确性方面也存在一定问题。据不完全统计,在已出版文献的引文数据中,存在不同程度的错误(约1%),如弄错论文题目或作者姓名中的符号,缺少介词,年代、卷期号、页码,或DOI等书写错误。为降低错误引文数据的数量,OCC开始借助外部资源提供的权威书目记录来纠正引文数据库中的错误,同时也提供引文数据在线检错服务<sup>[18]</sup>。

### 3.3 通过SPAR本体增强引文数据的语义性

传统的引文分析法只考虑在题录数据中列出的参考文献,关注文献间存在的引用关系,缺乏对引用行为的语义描述,包括引用功能(引文对施引文献的作用,如背景、基础、比较等)、观点倾向(否定或批判性引用会降低引文的价值)、引用频率(文献被引频次越高,其学术贡献越大)、引用位置(在方法、实验、结论等部分的引文通常比引言、研究现状或背景等部分的引文对于施引文献的学术价值更大)、施引文献的类型(如研究论文、技术论文、理论概念、案例分析、文献综述或一般性评论,因类型不同,其引用的价值也不同)等,这使得引用文献对于施引文献的具体贡献以及重要性无法准确展现,引用文献的相关信息如标题、关键词、摘要、作者、机构等不能被正确、全面地理解,而造成不同知识源间融合和互联的困难<sup>[10,19,20,21,22]</sup>。

针对上述问题,OCC通过引入描述学术交流、出版和引用信息的结构化领域本体SPAR,实现文献元数据和引文数据的自动化处理与互操作。如表1所示,SPAR由8个(FaBiO、CiTO、BiRO、C<sub>4</sub>O、DoCO、PRO、PSO和PWO)覆盖整个学术出版过程的通用本体构成,基本整合了出版与引用相关的所有属性。前4个本体主要用于描述书目对象、书目记录、参考文献列表、引用角色、引文记录、引文背景等;后4个本体主要为组织文档内容组件、出版角色、出版状态及出版流程提供结构化的控制词表。8个本体既可单独使用,也可以联成一体并用,各本体通过OWL2.0进行编码<sup>[22]</sup>。其中,CiTO

主要通过RDF语句中嵌入修辞性质的元数据来描述科技文献中不同参考文献的引用角色。实现该过程的工具为CiTO参考注释工具,其支持以自然语言和受控术语两种方式描述参考文献的语义特征<sup>[23]</sup>。

表1 SPAR中8个本体的功能

名称	功能
FaBiO	面向FRBR的书目本体,描述图书和期刊论文的书目信息
CiTO	包含不同引用关系及类型定义的引用本体
BiRO	反映参考文献题录特征的本体
C <sub>4</sub> O	反映文内引用特征的本体
DoCO	描述章、节、图、表等文档结构部件的本体
PRO	与出版功能相关的出版角色本体,描述学术出版过程各环节的关系
PSO	出版状态本体
PWO	出版流程本体

## 4 对我国引文数据开放关联的启示

从上述对OCC的分析看出,实现引文数据开放关联的关键点主要有以下内容。第一,尽可能争取出版商的支持,通过签署收割协议或者倡导开放获取行动为抓取书目资源及相关引文数据提供法律保障;同时,采用CC协议发布关联化后的引文数据,支持用户对引文关联数据的自由使用,实现创建OCC的初衷。第二,构建元数据模型,筛选核心实体,确定主要属性元素,明确不同实体间的关联关系,并为每个实体赋予唯一命名域,制定实例的命名规则,这是抓取实例数据的前提。通过元数据模型搭建书目资源内容表达、参考文献、载体表现、唯一标识符、代理及代理角色等实体(在实际应用中,根据具体需求,也可以扩展机构、项目、相关数据集等其他实体)的关联框架。第三,通过自动化程序提高实例数据的抓取效率,既包括从XML文档中抽取尽可能多的引文数据,转化为JSON格式存储,从源头上把大量非结构化信息变成结构化信息;也包括充分利用DOI、ISSN、ISBN、ORCID、URL或Crossref member URL等唯一标识符尽可能扩展引文信息,并用元数据元素作为谓词揭示这些信息与书目资源的各种语义关联,为后续知识推理及知识发现奠定基础。第四,利用SPAR本体为引文信息添加引用类型、引用角色等语义标签,实现对引用单元的语义标注,为用户理解引用内涵及科学引文分析奠定基础。第五,建

立与相关人员、机构、项目、事件、知识资源、数据库集等外部关联数据集的关联,多维度展示相关领域的隐性知识体系结构,丰富语义内容,实例间的关联关系通过类属性来揭示。

总之,OCC融汇了对书目资源内外部特征、引文数据的挖掘以及多维度知识关联的构建,同时关联了外部数据集,形成“从来源数据抓取书目及引文数据—扩展元数据—语义标注—知识关联”的工作流程,并以此为基础实现引文数据的开放共享。

我国机构数据库、学科数据库数量众多,也制定了相应的引文数据开放获取管理政策,但对于如何实施引文数据的开放获取计划,如何与出版商合作促进引文数据的开放与语义关联,还有待进一步探究。我国的图书情报服务机构应抓住机遇,制定可动态调整的引文分析发展战略,整合专业机构技术力量,与有关图情机构开展合作研究;同时,还要与计算机、网络技术人员合作,积极探索将引文数据发布为关联数据的方式方法,为我国引文数据的开放关联提供参考。

## 参考文献

- [1] ROUSSEAU R,LIU Y.Interestingness and the essence of citation[J].*Journal of Documentation*,2013,69(4):580-589.
- [2] CHADEGANI A A,SALEHI H,YUNUS M M,et al.A comparison between two main academic literature collections: Web of Science and Scopus Databases[J].*Asian Social Science*,2013,9(5):18-26.
- [3] National Steering Committee on Open Access Policy.National principles for open access policy statement[J].*Growth Hormone & Igf Research*,2015,25(1):28-33.
- [4] SWAN A.The open access citation advantage: studies and results to date[J].*Journal of Geophysical Research Atmospheres*,2010,112(FO2S06):195-225.
- [5] PIWOWAR H A,DAY R S,FRIDSMA D B.Sharing detailed research data is associated with increased citation rate[J].*Plos One*,2007,2(3):e308.
- [6] PIWOWAR H A,VISION T J.Data reuse and the open data citation advantage[J].*Peerj*,2013,1(3):e175.
- [7] Dalmeeth Singh Chawla.Now free:citation data from 14 million papers, and more might come[EB/OL].[2017-04-06].<http://www.sciencemag.org/news/2017/04/now-free-citation-data-14-million-papers-and-more-might-come>.
- [8] Open Citation Corpus(OCC)[EB/OL].[2017-04-06].<http://opencitations.org>.
- [9] Dryad Digital Repository[EB/OL].[2017-04-06].<http://datadryad.org/>.
- [10] 林芳芳,赵辉.美国Dryad数据库共享政策及启示[J].*中国科技资源导刊*,2015,47(6):48-52,94.
- [11] PERONI S, DUTTON A,GRAY T, et al.Setting our bibliographic references free: towards open citation data[J].*Journal of Documentation*,2015,71(2):253-277.
- [12] CrossRef metadata best practice to support key performance indicators(KPIs)for funding agencies[EB/OL].[2016-11-05].[http://fundref.crossref.org/docs/funder\\_kpi\\_metadata\\_best\\_practice.html](http://fundref.crossref.org/docs/funder_kpi_metadata_best_practice.html).
- [13] SHOTTON D.Open citations[J].*Nature*,2013,502(7471):295-297.
- [14] PERONI S,SHOTTON D.Metadata for the OpenCitations Corpus [EB/OL].(2016-07-07)[2016-11-05].<https://dx.doi.org/10.6084/m9.figshare.3443876>.
- [15] FALCO R,GANGEMI A,SILVIO P,et al.Modelling OWL ontologies with Graffoo[C]//In the Semantic Web:ESWC 2014 Satellite Events.Springer,2014,8798:320-325.
- [16] PERONI S,SHOTTON D,VITALI F.Freedom for bibliographic references:OpenCitations arise[EB/OL].[2016-11-05].<https://w3id.org/oc/paper/occ-lisc2016.html>.
- [17] FRANCESCHET M.A comparison of bibliometric indicators for computer science scholars and journals on Web of Science and Google Scholar[J].*Scientometrics*,2010,83(1):243-258.
- [18] SHOTTON D.Semantic publishing:the coming revolution in scientific journal publishing[J].*Learned Publishing*,2009,22(2):85-94.
- [19] PRIEM J.Scholarship:Beyond the paper[J].*Nature*,2013,495(7442):437-440.
- [20] CIANCARINI P,IORIO A D,NUZZOLESE A G, et al. Characterising citations in scholarly articles:an experiment[C]//International Workshop on Artificial Intelligence and Cognition.[S.l.]:[s.n],2013.
- [21] TEUFEL S,SIDDHARTHAN A,DAN T.An annotation scheme for citation function[C]//Proceedings of Sigdial Workshop on Discourse & Dialogue.[S.l.]:[s.n],2006.
- [22] BERGSTROM C T,WEST J D, WISEMAN M A.The eigenfactor (TM)metrics[J].*Journal of Neuroscience the Official Journal of the Society for Neuroscience*,2008,28(45):11433-11434.
- [23] PERONI S, SHOTTON D. FaBiO and CiTO: ontologies for describing bibliographic resources and citations[J].*Web Semantics Science Services & Agents on the World Wide Web*,2012,17(17):33-43.

## 作者简介

宋丹辉，女，1983年生，博士，馆员，研究方向：知识组织、知识服务，E-mail: hkdsongdh@163.com。

### The Open Citation Corpus Practice Case and Its Enlightenment

SONG DanHui

(Library of Henan University of Science and Technology, Luoyang 471023, China)

Abstract: As a new kind of linked data, the open citation data has already made substantial progress abroad, but has not yet to start at home. To shorten the gap, this paper chooses Open Citation Project which has played an important role in academic communication system as learning case. The author not only analysis the construction achievements and practices of its scale, technical route, the metadata model and working process in-depth, but also elaborates the methods and features in solving the problems of not easy to obtain, inconsistent, and semantic absence in citation dates. Moreover, summarizes the lessons to learn in aspects of development idea, key procedures, in order to provide sightedness suggestions to the construction of linked open citation data at home.

Keywords: Citation Information; Openness and Relevance; Open Citations Corpus

(收稿日期: 2017-04-08)

## ■ 书 讯 ■

# 《科技报告体系构建研究》

为推进我国科技报告制度建设，强化科技报告资源共享服务，贺德方研究员率领中国科学技术信息研究所科技报告研究团队，进行了国家自然科学基金重点项目“中国科技报告资源体系构建”（11ATQ006）研究，并对20多年来中国科学技术信息研究所相关研究和实践进行了归纳、凝练、整理和补充，最终形成了《科技报告体系构建研究》。

本书作为国家自然科学基金重点项目的主要研究成果，总结了科技报告产生发展的管理历程、凝练了科技报告制度的建设路径、制订了科技报告资源的整合方案，提出了科技报告体系的构建模式，归纳了科技报告实践的操作过程。本书对各级科技计划管理人员强化科技计划项目过程管理具有借鉴作用，对科研人员撰写高质量科技报告具有指导作用，对各类科研机构做好科技报告呈交、推进科技项目的规范管理和机构知识库建设具有参考价值，对图书信息机构做好科技报告深层次加工和收藏利用具有引导作用，也可供高校信息管理、科技政策与管理等专业研究生学习参考。

《科技报告体系构建研究》于2014年12月由科学技术文献出版社出版，定价78.00元。