

基于元数据的数据论文与期刊论文的 关联研究*

——以全球变化科学研究数据出版系统为例

陈帅印¹ 刘桂锋^{1,2} 刘琼²

(1. 江苏大学科技信息研究所, 镇江 212013; 2. 江苏大学图书馆, 镇江 212013)

摘要: 科研范式的不断转变以及科研人员数据意识的增强使得科学数据的价值逐渐升高, 而关联研究是为了研究不同知识、不同数据之间的相互关系, 发现潜在规律。探讨数据论文与期刊论文的关联应用能够推动科学数据高效流通、提高数据的应用能力。本文基于元数据, 以全球变化科学研究数据出版系统为数据来源, 通过对该平台的数据论文及相应的期刊论文进行关联的特征分析, 并选取有代表性的论文数据为研究对象, 进行二者关联应用模式的研究。

关键词: 元数据; 数据论文; 期刊论文; 关联研究; 数据管理; 科学数据

中图分类号: G250 **DOI:** 10.3772/j.issn.1673-2286.2022.08.002

引文格式: 陈帅印, 刘桂锋, 刘琼. 基于元数据的数据论文与期刊论文的关联研究——以全球变化科学研究数据出版系统为例[J]. 数字图书馆论坛, 2022 (8): 11-18.

科学研究第四范式(数据密集型科学)^[1]的产生使得科研人员将关注的焦点转向科学数据的组织、存储、开放、共享与使用等方面。科技文献的出版是研究人员发布科研成果的重要途径之一, 而在科研过程中产生的科学数据有十分重要的意义。为使科学研究过程可以得到重现的同时保护数据工作者的知识产权, 一种可将数据以特定的形式“出版”, 基于数据论文(data paper)概念的新数据出版模式开始被广泛认识并认可^[2-3], 随后出现了同时发表学术论文和数据论文的重叠期刊(overlay journal)、纯数据期刊(pure data journal)^[4]等新的出版物形式。数据论文率先出现在生物学、地球科学等领域^[5]。早在2000年, 美国生态学会(Ecological Society of America, ESA)^[6]就在投稿须知中对数据论文进行阐释, 认为它是一种特殊类型的论文, 用于展示大型或丰富的数据集, 包括描述数据内容、数据产生背景、数据质量和结构的元数据文件。

随着科学数据对科学研究重要性的不断显著, 数据论文的作用与价值也在学术出版中日益凸显, 不少学者对其概念、形式与结构、出版流程、作用等进行了研究: Candela等^[7]认为数据论文的概念应至少包含两个具体的信息对象——数据集(数据论文的描述对象)和数据论文本身(描述数据集的作品), 其中数据论文一词特指后者, 数据论文与传统期刊文章相同, 包括标题、作者、摘要、章节和参考文献等内容; 屈宝强等^[8]详细地对数据论文的形式、内容与结构进行阐述, 并且总结数据论文的出现能够强化数据利用、促进数据价值升值; Roa-Martínez等^[9]在Candela对15家出版平台的数据期刊的投稿指南、写作指导等板块进行分析后, 进一步提出了一种通用的数据论文构成体例。总的来说, 数据论文是对科研过程中数据收集、数据内容、数据处理、数据格式等进行详细描述, 类似于传统期刊论文中的方法部分, 但不提供任何数据的分析, 也没有产生新的研

* 本研究得到江苏大学科研课题立项项目“基于元数据的学术论文与数据论文的关联模式与设计研究”(编号: 21C312)资助。

究结论,其内容只是与科研过程产生的数据有关,可以在数据制作完成后迅速发表^[10-12]。

科学数据与期刊文献是科学研究的重要学术成果,但二者通常是相互独立地支持科学研究。若对数据与论文的关联应用进行研究,探索其中的关联特征与关联模式,揭示论文与数据之间清晰的关联性,可以更好地发挥二者的科学研究价值,更好地支撑数据高效应用,对于科学数据的获取与共享、科学数据的再利用和科研创新等有着重要的意义。部分学者在探讨二者关联问题时,在科学数据的元数据质量改善以及科学数据与期刊论文的关联模式方面的研究缺乏实证、应用层面的探索。因此,本文以元数据理论为基础,从科学数据生命周期理论、利益相关者理论等出发,聚焦于探索数据论文与期刊论文的关联模式,以实证应用为侧重点,运用有代表性的具体案例,将理论研究应用到实践中,切实实现科学数据能够在科研成果中的可应用、可共享和可推广的目标,推动科学数据高效流通、提高数据的应用能力。

1 科学数据与文献的关联研究

已有研究从多个角度对科学数据与文献的关联方式进行分析与总结。例如:卫军朝^[13]以Elsevier出版集团的科学数据与文献关联实践作为研究对象,总结归纳科学数据与文献的4种关联方式为可验证数据DOI、文本中的语义实体链接、数据仓储的横幅标识和集成数据App;姜恩波等^[14]描述文献与科学数据分离的现状,并通过案例介绍科学数据与文献关联融合的呈现方式,把科学数据与文献的关联方式分为基于形式的硬关联、基于内容的软关联、基于语义的融合。

本研究从内容、结构和目的等方面对相关研究进行归纳,将这些研究分为两大类:基于引文的科学数据与文献的关联,基于元数据的科学数据与文献的关联。

1.1 基于引文的科学数据与文献的关联

基于引文的科学数据与文献的关联研究主要以基于唯一标识符的关联和基于引文分析方法对参考文献等引用与被引用现象进行研究两大方面为主。

唯一标识符具有用于科学数据、学术文献等研究成果存储以及发布时的引用功能,主要包括数字对象唯一标识符(Digital Object Unique Identifier,

DOI)、国际标准关联(International Standard Link Identifier, ISLI)、统一资源名称(Uniform Resource Name, URN)等。标识符的使用能够以可引用、可发现的方式使更多的数字资源、科学数据可见、可用和可搜索。孙文佳等^[15]研究科学数据与科学文献的关联模式,并探讨了DOI在实现科学数据与科技文献关联研究中的关键作用;朱江等^[16]在《基于ISLI标准的科技文献和科学数据的关联》中,基于ISLI对科技文献和科学数据关联的ISLI元数据集扩展、编码体系设计等方面进行研究;Kraft等^[17]基于DOI对数字资源和文献的关联进行研究并探索永久识别码(Persistent Identifiers, PID)服务在学术内容与数据关联关系研究中的作用。

引文分析是对研究对象主要是学术文献、期刊文献、著者等的引用与被引用现象进行分析与研究,是学者进行关联研究的重要方式之一。科技文献引用科学数据是指科研人员在论文写作中引用已有的科学数据作为研究资料进行进一步研究,或利用已有的科学数据来证明自己的研究成果等。郭学武^[18]基于引文进行了科学数据与科技文献的关联研究,将科学数据与科技文献的关联模式分为3种:基于引用的直接关联、基于同被引的关联、基于引文的扩展关联;孙薇^[19]使用引文分析等方法探索科技文献关联数据的科研关系发现模型;张鑫等^[20]设计一种基于引文探针的关联算法,在高能物理领域文献与粒子关联场景下实现该算法,通过对关联度的计算发现更多的隐含关联,提高关联的覆盖率。

1.2 基于元数据的科学数据与文献的关联

科学数据的元数据以科学数据为描述对象,是对科学数据开展描述、组织、出版等工作的重要工具。科技文献的元数据以文献特征为描述对象,对文献进行知识组织、结构描述等,不同文献有着不同的元数据元素。基于元数据的科学数据与文献的关联侧重通过对科学数据的元数据与文献的元数据进行外部特征、内容特征的对比,并对二者的相关性、关联性进行研究。黄筱瑾^[21]从科学数据和科技文献的元数据出发,从两者的元数据描述中提取表达内容特征的元数据项,并利用向量空间模型进行特征的相似性计算,从而关联科学数据与科技文献;韩涛^[22]以生物信息学为例,对生物信息学科学数据库与文献库的外部关联和内部关联两层关系进行深度探究,基于内外部关联关系的交叉

引用, 对内容层面和知识层面进行聚合、关联。

综上可知, 大多相关研究从不同角度对科学数据与文献的关联进行分析, 已有一定的理论基础, 而基于引文的关联和基于元数据的关联之间基本上不会重叠, 只有在一定程度上存在交叉关系。基于引文的关联主要通过唯一标识符和参考文献两个方面进行探析, 而元数据元素是对科学数据与文献关联关系研究的重要指标之一, 元数据元素根据各个研究内容及研究结构的不同也会有不同的元数据元素分类, 当参考文献或者唯一标识符作为研究的元数据元素时, 则两者之间必然存在一定的交叉关系。本研究在前期提取数据时发现, 参考文献是元数据元素之一, 这也表明基于引文的关联与基于元数据的关联在本研究不是并列关系, 而是存在一定的相互交叉融合的关系。

2 数据论文与期刊论文关联特征研究

2.1 数据来源

全球变化科学研究数据出版系统是中国科学院地理科学与资源研究所、中国地理学会主办, 以全球变化科学研究数据出版为核心的出版系统。该平台从2014年开始收录数据论文, 每年期数不等, 每期有10~20篇数据论文出版。截至2021年第12期, 全球变化科学研究数据出版系统平台共有50期、数据论文总数量1019篇。

“关联论文”字段是全球变化科学研究数据出版系统平台数据论文的元数据元素之一, 通常指利用数据集完成的研究成果, 该字段主要使读者能够快速检索并阅读与数据论文相关联的期刊论文。因此, 本研究以全球变化科学研究数据出版系统为数据来源, 对其数据论文及关联论文字段进行研究, 以探索数据论文和期刊论文的关联关系。

从全球变化科学研究数据出版系统人工手动采集2016年第1期至2021年第12期发表的全部存在“关联论文”字段的数据论文相关数据, 包括数据论文题目、数据集贡献者/作者、数据论文作者单位、数据论文关键词、关联论文篇数、关联论文题目、关联论文作者、关联论文类型等, 并通过“关联论文”字段检索到期刊论文, 并采集相关数据, 包括期刊论文题名、作者、作者单位、关键词等。数据收集时间截至2022年1月15日, 收集数据论文总量938篇, 其中拥有“关联论文”字段的数据论文共有389篇。

2.2 数据论文文献特征分析

与期刊论文一样, 数据论文的文献特征也可以划分为外部特征和内容特征^[23], 具体见表1所示, 可以看出两者是一一对应的关系。

表1 数据论文与期刊论文的外部特征与内容特征

特 征	数据论文	期刊论文
外部特征	数据集贡献者/作者	文献作者
	数据集作者所属单位	文献作者所属单位
	出版时间	文献发表日期
	基金项目	基金项目
	数据下载	文献下载
	数据引用方式	文献引用方式
内容特征	数据集名称	文献题名
	数据集摘要	文献摘要
	关键词	关键词

2.3 数据论文与期刊论文关联情况分析

从数据论文及其关联的期刊论文的发文数量及趋势来看, 2018年的关联论文总量占数据论文总量比例最大, 达63.10%, 而2019年、2020年占比分别为32.89%、32.53%, 原因在于2019—2021年该平台出版大量属于“全球变化数据大百科辞条”类型的数据论文, 导致关联论文的占比略有下降, 但并不影响数据论文的发表数量及发表进展。

从关联论文的文献类型看, 大部分关联论文的类型在2016年为“中文期刊”, 2017年开始, 关联论文的文献类型不再仅限于中文期刊、英文期刊, 增加了“年度报告”, 此后文献类型逐渐丰富, 发展至2021年, 关联论文类型呈现“中文期刊”“英文期刊”“专著”“年度报告”“学位论文”“论文集”等多样化趋势, 这意味着二者的关联度更加紧密, 而这样的现象反映出科学数据在期刊论文中的应用是多元化的。结合文献类型的数量统计及每年的变化趋势不难看出, 数据论文与其关联论文之间不再单一化, 与数据论文对应的关联论文数量每年都有所改变, 但文献类型随着时间的推进而逐渐多样化, 因此形成了从原来的一篇数据论文仅有一篇关联论文发展到一篇数据论文可以关联到多篇期刊论文的现状, 并且与数据论文相关的关联论文的文献类型呈现多样化趋势。

从数据论文及其关联期刊论文的作者来看,大部分都有合作关系。究其原因,在大数据时代背景下,大部分科学研究依赖实验、数据模拟或观测等研究方法产生的海量数据(集)^[24],而单个学者很难在兼顾科学研究的同时完成科学数据的采集、处理、加工、分析等全部工作,进行作者合作完成数据共享已经成为科学数据共享的主要形式。合作关系主要有弱合作、一般合作、强合作等层次的区分,强合作主要涵盖的数据论文与其关联密切的期刊论文的作者群一模一样,弱合作主要指第一作者或者第一、第二作者相同。

3 基于元数据的数据论文与期刊论文关联模式研究

元数据是指提供关于信息资源或数据的一种结构化的数据,是对信息资源结构化的描述,其作用是描述信息资源或数据本身的特征和属性。本研究对元数据描述资源框架的设计主要参考了DC元数据标准。通过对基本元素的复用,可以将元数据元素分为期刊论文元素修饰词和数据论文元素修饰词两大项(见表2)。

表2 期刊论文元数据元素与数据论文元数据元素

元 素	元素名称	期刊论文元素修饰词	数据论文元素修饰词	复用标准/自定义
Title	题名	论文题名	数据集名称	DC
Date	日期	接受日期	出版日期	DC
		修改日期		
		发布日期		
Creator	主要责任者	责任方式	数据集贡献者	DC
		机构	数据集创建者	
Subject	主题	关键词	关键词	DC
Publisher	出版者	期刊单位	发布机构	DC
Type	资源类型	-	-	DC
Description	描述	目录	数据集描述	DC
		摘要		
		基金		
Contributor	其他责任者	责任方式	-	DC
		机构		
Format	格式	范围	-	DC
		媒体		
Source	来源	-	-	DC
Rights	权限			
Identifier	标识符	文献引用	引用方式	DC
Language	语种	-	-	DC
Relation	相关资源	参考文献	参考文献	DC
		译文	关联论文	
Coverage	-	-	-	-

*注:“-”表示无此项

对数据论文、期刊论文的元数据元素进行修饰后,可以从单个元数据元素进行相对应的关联应用研究,从而探究数据论文与期刊论文基于元数据元素的关联模式。具体而言,期刊论文与数据论文的关联可以通过研究二者的元数据元素之间的关联应用使得二者的关联模式更加清晰,进而实现基于元数据的关联。鉴于此

构建基于元数据元素的关联应用机制(见图1),从数据集名称、数据集贡献者/作者、贡献者/作者单位、出版时间等多个数据论文的元数据元素展开,与论文题名、论文作者、所属单位、论文发表时间等期刊论文的元数据元素一一相照应,对二者的元数据元素进行关联应用的深度探讨。

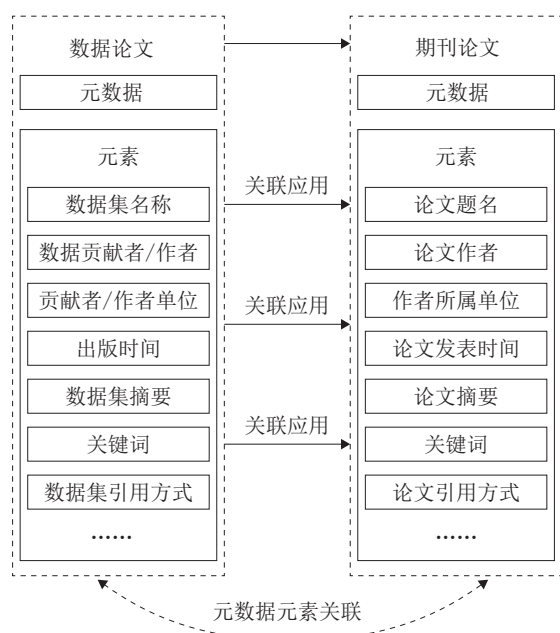


图1 基于元数据元素的关联应用机制

3.1 基于作者/作者单位的关联模式

对于实验、观测等科研过程中产生的数据，数据生产者也就是研究人员在完成科技文献的撰写后，还会将这些数据整理成数据论文发表在数据期刊上，成为学术研究成果的一部分，那么该数据论文和期刊论文基于元数据或者其他因素都存在一定层面的关联。

一般情况下，一项科学研究不仅靠一个研究人员就能完成，因此存在跨学科、跨领域等方式的科研合作关系，恰是作者合作关系网将本没有关系的数据论文与科技文献关联起来。基于作者合作关系网的数据论文与期刊论文的关联主要分为3种类型：不同作者或者研究团队共同完成数据论文的撰写，并且该数据论文与不同作者各自完成的期刊论文构成直接关联；不同作者合作完成的期刊论文与其中一个作者撰写的的数据论文构成间接关联；同一作者完成的不同数据论文之间构成互引式关联。

3.2 基于内容特征的关联模式

数据论文和期刊论文一样有外部特征和内容特征，内容特征主要包括论文（数据集）名称、论文（数据集）摘要、关键词。数据论文和期刊论文的元数据元素的修饰词虽然有一定的差异，但其描述的元数据元素本质内容是相同的，尤其在内容特征元素的描述上

存在相互映射的联系。通过提取二者的内容特征进行关联应用的研究，可以从内容特征角度进行内容之间的关联，从而使得二者的关联模式更加明晰。

数据论文的论文（数据集）名称与其相关联的期刊论文的题名存在一定的映射关系，期刊论文中使用的科学数据主要名称大多会出现在数据论文的论文（数据集）名称中。

期刊论文的摘要主要用于对整篇文章的内容、研究方法、研究结论等进行描述，能够使读者在阅读完摘要后就能对整篇文章的结构有一个较为清楚的认识。数据论文的摘要内容没有过多对研究方法、研究流程的描述，大多会详细地阐述该数据集的来源、用途、数据量、数据格式等。在常见的对科学数据进行统计计量与分析的文章中，可以明显发现在摘要部分会提及数据的来源渠道等，鉴于此，二者的摘要部分有一定的关联应用。

期刊论文的关键词一般包含研究对象、研究方法、研究主旨等，读者通过关键词能够确定论文的研究主题等，对阅读原文有一定的帮助。在数据论文中，关键词主要涉及数据集的名称、数据处理与分析工具、数据集指标等，这与期刊论文的关键词中有部分内容的映射和重合。在使用科学数据的期刊论文中，部分论文会以数据集的研究对象名称、数据集部分指标等作为该论文的部分关键词，其中有着基于内容的关联应用。

3.3 基于引用方式的关联模式

引用行为是指在各种因素影响下，论文作者在撰写相关的论文时参考或者借鉴其他学者的学术文献并加以标注说明的学术行为，无论是期刊论文还是数据论文都存在引用行为。数据论文的参考文献主要涵盖期刊文献、会议报告等，科学数据集的形成是多源的，同一研究对象的数据在不同文献类型出现，作为数据论文需要将这些数据统一收集并出版。而对于期刊论文来说，参考文献大多是其他学者的学术文献等，但也不排除有部分期刊论文在撰写过程中会用到数据论文或者数据集的情况。基于引用方式角度进行数据论文的参考文献以及期刊论文参考文献之间的关联应用研究，可以更好地为二者的关联模式提供参考。

引用位置指引用内容在施引文献中出现的位置，从引用位置角度进行分析，能对引文出现位置的不同性与相似性进行概述，为关联研究提供思考的新角度。

3.4 基于“关联论文”字段的关联模式

“关联论文”字段是全球变化科学研究数据出版系统数据论文的元数据之一，该字段可以为阅读者提供更便捷的方式去阅读相对应有数据关联的论文。该关联模式以数据论文和期刊论文之间是否有数据应用关系为主要指标，涵盖二者间的各类元数据元素、论文类型、论文内容等，“关联论文”字段的论文类型涉及多种

文献形式，包含中文期刊、英文期刊、数据论文、年度报告等。使用本体编辑和知识获取软件Protégé对此类数据进行本体构建，并对“期刊论文”和“数据论文”两本体间构建关联关系为“关联论文”（见图2）。二者的属性分布主要为论文的外部特征和内容特征，看似无关联的期刊论文和数据论文，在全球变化科学研究数据出版系统上通过“关联论文”字段有了内容、元数据等方面的关联应用。

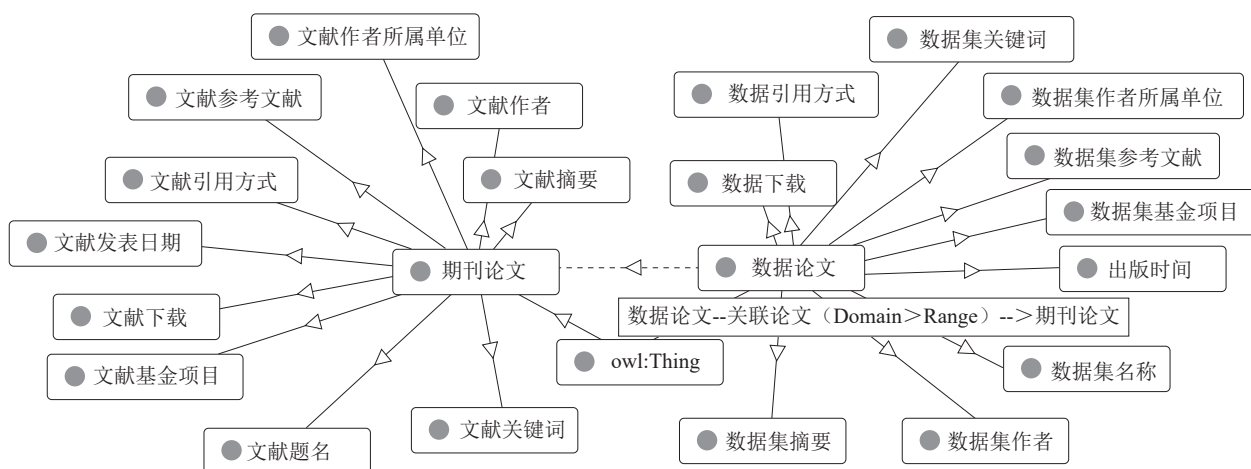


图2 数据论文与期刊论文关联的本体构建与可视化

为验证上述提出的期刊论文与数据论文应用关联模式的适用性，遵循随机抽样性的原则从2020年、2021年的数据集中抽取含有“关联论文”字段的数据论文及相对应的期刊论文，并采集二者数据集的基本信

息、数据集应用与处理的软件或者工具、数据集应用在期刊论文中的位置、数据集应用目的等数据，进行期刊论文与数据论文的应用关联分析（见图3）。

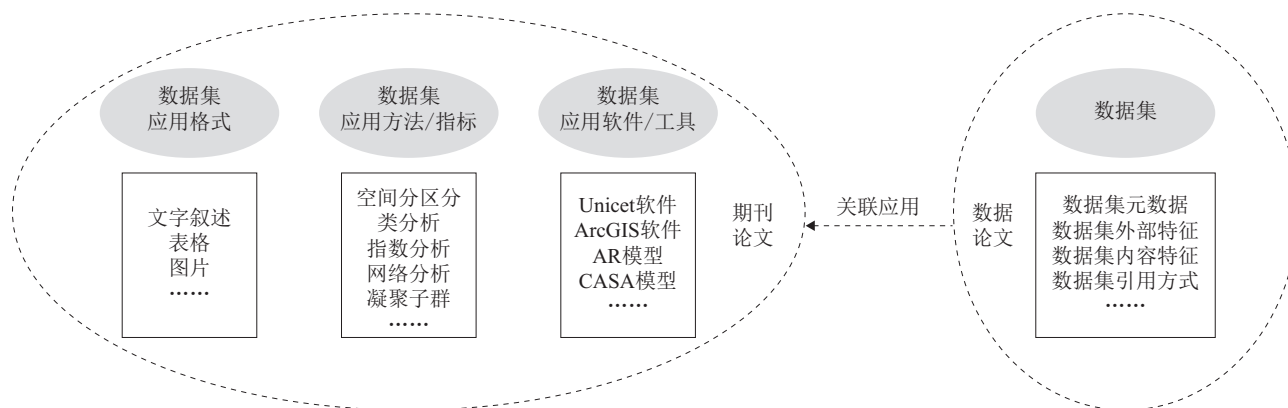


图3 基于数据集应用的数据论文与期刊论文关联

数据集的应用格式主要以文字叙述、表格、图片3种类型为主，由于全球变化科学研究数据出版系统的学科类型偏向地学，数据集在期刊论文中的应用格式

绝大部分以图片为主，在图片的前后文会以文字形式进行简单的概述和介绍。数据论文对数据集的元数据、外部特征、内容特征、引用方式等多个维度进行描述。

科学研究从收集数据、处理数据、分析数据到最终数据呈现形式的过程中,通常使用多种应用方法或指标,通过不同的软件、工具将数据转化。该平台数据集的应用方法/指标与其学科密切相关,一般运用到空间分区分类分析、各类指数分析、网络分析、凝聚子群等多种与地学相关的研究方法或指标,采用的软件/工具主要以Unicet、ArcGIS等为主,AR模型或者其他建模工具为辅,用于将前期的科学数据可视化。因此,数据论文描述的数据集实际应用于期刊论文中,由数据集的应用格式、方法或指标、使用的软件或工具等架构二者之间的关联。

除了对数据集的应用格式、方法或指标、软件工具等进行研究,还可以对数据集的应用位置及数据集的应用目的等进行多个方向的阐述与分析。数据集的应用位置主要出现在期刊论文的研究背景、数据来源、结果分析等部分中,即论文中使用数据集的位置大多是研究的主要内容部分。研究背景部分是介绍研究的大背景及研究来源,数据集在此部分的应用可以印证该研究的意义、佐证研究的可行性。数据来源部分是研究性论文的数据阐述部分,数据集在此部分的应用能够直观表述数据源。结果分析是较多学者在对数据进行采集、处理、分析后表述分析结果的部分,数据集应用于结果分析部分为研究结论提供数据支撑,由此可见科学数据在一项科学研究过程中的重要性。

科学数据通过前期大量的数据采集、中期的处理、后期的分析,其最终的研究目标是解决学术问题或未解决的实际问题,通过科学研究从而解决该问题,彰显科学学术意义、为社会创造价值,因此数据的价值便在数据集的应用目的这一方面得到凸显。在变化科学研究数据出版系统中,数据集在期刊论文中的应用目的主要为构建指数序列、呈现时空特征、构建参数、分析时空格局演变、分析影响因素等。

4 结论与展望

关联研究是研究不同知识、不同数据之间的相互关系,并从中发现研究对象之间的潜在关系^[25]。将海量科学数据以数据论文的形式发表从而起到数据共享的目的,基于元数据对数据论文与期刊论文的关联进行分析,是科学数据共享成为大势所趋的发展方向。本文以元数据为切入点,通过对全球变化科学研究数据出版系统的数据论文进行元数据相关项的提取,并通过“关

联论文”字段提取相关联的期刊论文元数据相关项,对二者进行关联融合研究。先从论文的文献特征、文献类型、发文量与趋势等方面对二者进行论文文献的计量与分析,接着选取代表性案例数据进行深度探讨,对论文进行基于元数据的关联应用研究,概述该平台的数据论文及关联的期刊论文关联模式主要为四大类:基于作者/作者单位的关联、基于内容特征的关联、基于引用方式的关联和基于“关联论文”字段的关联。本研究从数据的元数据、内容到数据的应用方式与应用效果逐步进行分析,侧重数据元素的应用研究。

在知识创新日益重要的今天,对科学数据的研究不再停滞于理论层次,而是要对科学数据的应用方式以及二次利用模式进行探究。期刊论文是学术成果展示的重要形式之一,对数据论文与期刊论文的关联进行揭示,将对科学研究和科技创新产生重要的推动作用。基于元数据、数据应用方式的关联研究只是数据与文献关联的完整研究体系中的一部分,随着科学的不断发展,各个领域的科学研究将不断产生新的科学数据,新的学术成果展示形式也会不断更新。因此,在今后的研究工作中还需从不同的视角对数据与文献进行探讨分析,促进科学数据与文献之间的深度融合,进而推进关联研究的深入发展。

参考文献

- [1] 储节旺,夏莉. 嵌入生命周期理论的科学数据管理体系构建研究——牛津大学为例[J]. 现代情报, 2020, 40(10): 34-42.
- [2] KUNZE J A, CRUSE P, HU R, et al. Practices, Trends, and Recommendations in Technical Appendix Usage for Selected DataIntensive Disciplines [EB/OL]. <https://escholarship.org/uc/item/9jw4964t>.
- [3] CHAVAN V, PENEV L. The data paper: a mechanism to incentivize data publishing in biodiversity science [J]. BMC Bioinformatics, 2011, 12(15): S2.
- [4] 伏安娜,张计龙,殷沈琴. 数据论文国内外发展研究综述[J]. 图书情报工作, 2015, 59(24): 131-138.
- [5] 马建玲,曹月珍,王思丽,等. 学术论文与科学数据集出版研究[J]. 情报资料工作, 2014(2): 82-86.
- [6] 邢文明,刘婷. 增强出版驱动的科学数据出版: 动因、模式及路径[J]. 中国科技期刊研究, 2019, 30(8): 853-861.
- [7] CANDELA L, CASTELLI D, MANGHI P, et al. Data journals: a survey [J/OL]. Journal of the Association for

- Information Science and Technology, 2015, 66 (9) [2022-08-01]. DOI:10.1002/asi.23358.
- [8] 屈宝强, 王凯. 数据论文的出现与发展 [J]. 图书与情报, 2015 (5): 1-8.
- [9] ROA-MARTÍNEZ S M, VIDOTTI S A B, SANTANA R C. Proposed structure of a data paper structure as scientific publication [J]. Revista Española de Documentación Científica, 2017, 40 (1) [2022-08-04]. DOI:10.3989/redc.2017.1.1375.
- [10] CALLAGHAN S, MURPHY F, TEDDS J, et al. Processes and procedures for data publication: a case study in the geosciences [J]. International Journal of Digital Curation, 2013, 8 (1): 193-203.
- [11] GORGOLEWSKI K J, MARGULIES D S, MILHAM M P. Making data sharing count: a publication-based solution [J/OL]. Frontiers in Neuroscience, 2013, 7 [2022-08-01]. https://doi.org/10.3389/fnins.2013.00009.
- [12] NEWMAN P, CORKE P I. Data papers-peer reviewed publication of high quality data sets [J]. The International Journal of Robotics Research, 2009, 28 (5): 587.
- [13] 卫军朝. 科学文献与科学数据关联实践研究——以Elsevier为例 [J]. 国家图书馆学报, 2017, 26 (3): 93-101.
- [14] 姜恩波, 裴玉香. 科学文献与科学数据的融合方法与实例研究 [J]. 知识管理论坛, 2019, 4 (2): 69-79.
- [15] 孙文佳, 常娥. 科学数据与科技文献关联分析 [J]. 图书馆理论与实践, 2017 (3): 49-53.
- [16] 朱江, 李欣怡, 姜恩波, 等. 基于ISLI标准的科技文献和科学数据的关联 [J]. 图书馆理论与实践, 2020 (5): 80-83.
- [17] KRAFT A, DREYER B, LÖWE P, et al. 14 years of PID services at the German National Library of Science and Technology (TIB): connected frameworks, research data and lessons learned from a national research library perspective [J/OL]. Data Science Journal, 2017, 16 (1/2) [2022-08-01]. DOI:10.5334/dsj-2017-036.
- [18] 郭学武. 基于引文的科学数据与科技文献关联研究 [J]. 情报科学, 2014, 32 (4): 59-62.
- [19] 孙薇. 基于科技文献关联数据的科研关系发现研究 [D]. 镇江: 江苏大学, 2017.
- [20] 张鑫, 文奕, 杨宁, 等. 基于引文探针的文献与数据的关联算法与应用——以高能物理领域为例 [J]. 情报理论与实践, 2019, 42 (10): 151-156.
- [21] 黄筱瑾. 基于内容特征的科学数据与科技文献关联研究 [J]. 现代情报, 2018, 38 (1): 56-59.
- [22] 韩涛. 科学数据与科学文献相关性研究——以生物信息学为例 [J]. 图书情报知识, 2008 (3): 42-46.
- [23] 秦春秀, 杨智娟, 赵捧未, 等. 面向科技文献知识表示的知识元本体模型 [J]. 图书情报工作, 2018, 62 (3): 94-103.
- [24] 丁大尉. 大数据时代的科学知识共生产: 内涵、特征与争议 [J]. 科学学研究, 2022, 40 (3): 393-400.
- [25] 陈明艳. 政府开放数据的语义描述与实体同一性研究 [D]. 广州: 华南理工大学, 2015.

作者简介

陈帅印, 女, 1998年生, 硕士研究生, 研究方向: 科学数据管理。

刘桂锋, 男, 1980年生, 博士, 研究馆员, 通信作者, 研究方向: 科研数据管理、大数据分析、专利分析, E-mail: liuguifeng29@163.com。

刘琼, 女, 1986年生, 硕士, 馆员, 研究方向: 阅读推广。

A Study of the Association of Metadata-based Data Papers with Journal Papers:
Take the Global Change Science Research Data Publishing System as an Example

CHEN ShuaiYin¹ LIU GuiFeng^{1,2} LIU Qiong²

(1. Institute of Science and Technology Information, Jiangsu University, Zhenjiang 212013, China; 2. Jiangsu University Library, Zhenjiang, 212013)

Abstract: The changing paradigm of scientific research and the increasing data awareness of researchers have gradually increased the value of scientific data, and the correlation study is to investigate the interrelationship and discover potential relationships between different knowledge and data. Exploring the application of association between data papers and journal articles can promote the efficient flow of scientific data and improve the application of data. Based on metadata, this paper uses the Global Change Scientific Research Data Publishing System as the data source, analyzes the characteristics of the data papers and the corresponding journal papers on the platform, and selects the representative paper data as the research object to study the application model of association.

Keywords: Metadata; Data Paper; Journal Paper; Association Research; Data Management; Scientific Data

(收稿日期: 2022-07-19)