

从关联科学国际研讨会看关联科学的研究与实践

刘磊, 郭诗云

(南京农业大学信息管理系, 南京 210095)

摘要: 随着关联数据技术的发展, 科学研究方式也从数字科研 (e-science) 向关联科学 (Linked Science) 转变。对关联科学国际研讨会进行分析, 旨在了解关联科学的发展历程与现状进展, 为国内关联科学研究提供参考与启发。通过对比分析了历届关联科学国际研讨会的召开宗旨与目标、研讨话题, 并从关联科学的应用与案例研究、关联科学数据的管理和检索、科学信息的定量与定性集成、本体技术在科学信息上的应用、数据语义化描述以及科学信息的溯源、质量和共享研究等6个方面对会议论文进行梳理与总结。分析得出国内研究还比较薄弱, 相关研究应进一步加强。

关键词: 关联科学; 科学数据; 语义网

中图分类号: G250

DOI: 10.3772/j.issn.1673-2286.2015.09.006

科学数据是科学研究过程的重要产出, 是科学研究与教育的基本信息资源。一般嵌入至论文、专著中, 包括各种数值型、事实型和文字型数据^[1]。随着计算机技术、网络技术和通信技术的发展, e-science 科学研究环境随之出现。e-science 是通过建立一种信息化的科研环境, 为科技工作者从事科学研究活动提供信息化的手段^[2], 其核心理念就是科学研究活动中借助先进的网格技术进行全球性合作^[3]。然而, 在e-science环境中, 科学数据的可重复性以及其共享模式一直难以得到有效实现。随着近些年语义网、本体、云计算、关联数据等技术的发展, 关联科学 (Linked Science) 由此产生。关联科学是一种实现科学资产互联的方法, 促进透明的、可重复的和跨学科的研究^[4]。

1 关联科学的提出

关联科学是由德国明斯特大学 (University of Munster) 地理信息科学研究所 (Institute for Geoinformatics) 的 Tomi Kauppinen 和巴西国家空间研究所

(National Institute for Space Research) 的 Giovana Mira de Espindola 于 2011年提出的。在2011年的国际计算科学大会 (International Conference on Computational Science) 上, 提交了一篇名为《关联开放科学: 交流、共享和评估可执行论文的数据、方法与结果》^[5]的论文。该论文指出, “Linked Science” 是 “Linked Open Science” 的简称, 说明了关联科学关联的主要内容是开放的科学资源和数据。关联科学的目的是为解决可执行文件面临的挑战而提出的一种方法, 是关联数据、语义网与Web标准、开放资源与网络在线环境、云计算、知识共享许可协议 (CC协议) 构建的技术与法律基础设施等的组合, 关联科学的形成包括了科学界和学术出版商的共同努力^[6]。

2 关联科学国际研讨会会议概述

2011年初, 德国明斯特大学的地理信息学研究所创建了LinkedScience.org网站。为推动关联科学的发展, LinkedScience.org自2011年起每年召开一届关联

科学国际研讨会 (International Workshop on Linked Science, 简称LISC), 至今已召开4届, 共产生33篇与关联科学相关的文献。这些论文主要从关联数据、语义网及本体、科学数据再利用等角度研究关联科学的相关问题。这4届研讨会分别是第10-13届国际语义网大会 (International Semantic Web Conference) 的专题研讨会之一。

3 关联科学国际研讨会的目标与意义

科研产出在传统上大多以文章的方式来公布。据统计, 全世界每年发表上百万篇科研论文, 仅在PubMed上平均每分钟就有一篇文章发表^[7]。验证科研结果需要可重复的研究方法, 其前提是原实验中使用的相同数据、过程和算法的可获取性。文章、方法和数据集是相互关联的, 但却不一定是可开放获取和相互链接的。因此, 数据的再利用是重要挑战。此外, 由于科研数据的大规模性和复杂性, 也使得数据的核实验证、再生共享和再利用变得愈发困难。语义网技术为实现科学数据的共享提供了一个有效方法, 通过对每个数据对象唯一标识并进行语义扩展, 极大地提高了数据的可获取性。关联科学的方法, 即发表、共享和互联科学资源与数据, 对科学研究至关重要, 是促进科研可再生和学科领域内及跨学科合作的关键。研讨会通过对关联科学的关键技术、实现方法、应用案例、问题与挑战等方面的探讨与交流, 一方面旨在促进科学研究者用关联科学的方法管理科学数据, 另一方面鼓励各领域技术专家协同合作, 促进跨学科研究。

LISC2011研讨会探讨了新兴技术 (关联数据、语义技术等) 能否或是如何实现关联科学^[7]; LISC2012研讨会重点探讨了大数据环境中的数据处理问题, 即如何利用语义技术和关联数据方法来解决大数据面临的挑战^[8]; LISC2013研讨会旨在讨论和展示使用语义网技术来发表、共享和互连科学数据的新方法的应用结果, 并运用这些资源来发现新的连接, 以验证、再利用和再生科学研究^[9]; LISC2014研讨会旨在“通过关联科学展现数据的意义”, 从而促进知识发现^[10]。

从历届关联科学研讨会的会议目标可以看出, 第一、二届研讨会重点在探索和尝试关联科学的实现技术和方法, 分析关联科学的必要性和可行性, 其中第二届会议基于当前的大数据环境, 集中讨论了大数据之间的关联问题。而第三、四届会议目标开始了从理论到

实践的转变, 重点关注关联科学方法在科学调查和实验中的应用问题, 并鼓励应用扩展, 说明在前两届研讨会后关联科学有了一定的理论成果。总体而言, 研讨会通过探索发表、共享、链接和分析现有科学资源结果的新方法, 并通过推理来发现新的关联和科学思考。

4 关联科学国际研讨会的话题分析

4届国际关联科学研讨会上的话题数量分别是19、22、18及24个。通过分析可知, 4届会议上共同出现的话题有9个, 说明这9个话题是关联科学领域重点关注的核心问题或一直尚未解决的问题。在LISC2011和LISC2012研讨会上, 出现较多的共同话题有9个, 说明前两届研讨会研究讨论的主题较为一致, 关联科学领域的研究重点并没有发生太大变化。其中LISC2011会上出现的“解决关联科学接受障碍的方法”在后两届上没有继续出现, 可见在此问题上专家学者已做出一定的研究。LISC2013和LISC2014研讨会中出现4个共同话题, 集中在语义网技术在关联科学中的应用及扩展, 说明经过前两届研讨会和一定量相关研究后, 关联科学有了新进展, 从基础理论研究向实践研究转变。

5 关联科学国际研讨会的会议文献分析

对4届关联科学国际研讨会的会议文献主题进行分析, 这些文献主要从关联数据、语义网及本体, 以及科学信息的来源、管理、检索和共享等角度研究关联科学的相关问题, 并在不同学科领域进行实践。文献研究的主题有6个方面, 具体包括关联科学的应用与案例研究、关联数据的管理和检索、科学信息的定量与定性集成、本体技术在科学信息上的应用、数据语义化描述以及科学信息的来源、质量和共享研究。

5.1 关联科学的应用与案例研究

P.Groth和Y.Gil针对网络科学 (Network Science) 中的关联数据进行研究。网络科学技术包括网络分析、网络建模和可视化, 而其关键难题就是数据获取问题。网络数据需要经常从非网络来源中挖掘和转换, 此过程通常需要耗费大量的时间精力并容易造成数据偏差, 数据的完整性也难以得到保障。针对这些问题, 作者提出了一种LinkedDataLens的实现方法, 从关联数

表1 各届关联科学研讨会话题分布

关联科学研讨会	话题 (Topics)	
四届研讨会 共有话题	• 关联科学的案例研究	• 科学信息的来源、质量、隐私和信任
	• 科学数据的格式化表示	• 科学数据集群来源的语义集成
	• 科学信息的定量与定性集成	• 科学应用的语义相似性
	• 本体技术在科学信息上的应用	• 基于共享和再利用的数据出版支持
	• 科学数据的本体可视化	
LISC2011、LISC2012 共同话题	• 基于关联数据的应用研究	• 法律、伦理和经济方面的关联数据研究
	• 科学出版物和潜在研究数据集的连接	• 传播和归档研究结果、合作和研究网络、 研究评定中的关联数据研究
	• 通过连接和数据集成的丰富语义数据	• 关联科学数据集的推理性机制
	• 研究成果的传播与归档、协作与研究网络、 研究评估的关联数据研究	• 语义驱动数据集成
	• 生命周期在关联科学的实现和提出解决策略	
LISC2013、LISC2014 共同话题	• 关联公民科学	• 关联数据集的科学推理机制
	• 基于关联数据的科学实验	• 科学信息检索
各届研讨会 单独话题	• 解决关联科学接受障碍的方法 (LISC2011)	• 关联模式研究 (LISC2014)
	• 大规模关联数据分析 (LISC2012)	• 语义查询生成 (LISC2014)
	• 科学信息的格式化编码 (LISC2012)	• 自动假设生成 (LISC2014)
	• 科学中大规模的数据集成 (LISC2012)	• 增强人类推理 (LISC2014)
	• 现存语义的可扩展性和管理科学资源的 关联数据解决方法 (LISC2012)	• 交互式语义系统 (LISC2014)
	• 使用语义网的科学资源传播与集成 (LISC2013)	• 关联数据实证研究的方法论探索 (LISC2014)
	• 激励复制别人的研究&使自己的研究重用 (LISC2013)	• 引言生成 (LISC2014)
	• 语义网实验的重用 (LISC2013)	• 使用语义Web的实验数据的整合 (LISC2014)
	• 语义网方法和结果的重用 (LISC2013)	• 实验步骤共享和再现研究 (LISC2014)
	• 使用语义网技术支持重用与再生 (LISC2013)	

据中提取网络关系并利用网络分析算法进行分析,其分析结果反过来又能以赋予特征的关联数据的形式发布^[11]。A.Di Iorio等提出一种语义镜的方法,用来加强期刊论文的语义关联性,并开发了原形TAL进行实证与应用^[12]。

S.Mäs等以GLUES项目研究结果为依据,探讨了一个以清晰和可追踪的方式在网络上发布科学地理数据所需的关联数据的方法,基于此建立了一个关联数据范式^[13]。V.Nguyen等面向生物医学领域,提出将科学著作知识和实验数据语义化集成,才能有效挖掘两种信息资源,更好地帮助研究人员理解潜在疾病的病理过程。他们提出了在两种知识来源间巩固和互补性使

用的语义方法,并在查加斯病研究上实验验证该方法的可行性^[14]; L.Kazemzadeh等通过关联数据技术开发了“LinkedPPI”框架来促进蛋白质互作(PPI)动力学领域的知识发现与应用^[15]; J.Schneider等研究了药物互作(PDDIs)领域的知识表示模型^[16]; S.Jupp等提出ReDrugS框架,提供了一个语义集成系统,结合不同数据结构的生物医学资源生成一个广泛的知识重用图谱^[17]。S.Nordhoff等介绍了Glottolog/Langdoc项目,该项目旨在提供近乎完全的关于世界语言的描述性资源书目范围,并且提供与其参照书目唯一的标识符,允许第三方来收获、消化和注解资源^[18]。A.Meroño-Peñuela等研究了关联人文科学数据,探索了语义网如何适用于人文数

据的方法,并以荷兰1795-1971年间的历史人口普查数据为研究对象进行实证,通过描述一种基于标准关联数据原则的统一协调方法,旨在说明生成的关联数据源的规模及复杂性如何向语义网技术提出新的挑战,并探讨了可能的解决方案^[19]。H.Ghaem Sigarchian等探讨了将视频、模型、幻灯片、数据集等多种形式的资源整合成数字版本形式的科学出版物的潜在可能性,并通过EPUB3标准在计算机领域进行实证^[20]。

以语义格式公布的生物学数据数量呈稳定性增长。特别是从数据库,如KEGG和WikiPathways,路径信息能以一种标准的从BioPAX转化的基于RDF格式进行获取。然而这些模型都是描述性的而非可执行性的。T.Willemsen等利用语义描述路径模型的可获取性,提出一种自动创造执行性路径模型的方法,并封装在一个名为BioPax2PNML的在线工具中。关联数据和语义网技术促进信息的合并和获取是实现该方法的关键基础^[21]。C.McLean等提出基于分子生物实验的设计模型方法,提供了一种可供选择的术语和解释框架,以此捕获专家经验和意图,这是目前利用网络本体和计算机工作流的实验室方法表述中所缺失的^[22]。N.Ockeloen等介绍了传记网(BiographyNet),旨在从各种荷兰传记词典中提取人、历史事件、地点和时期来创建一个语义知识基础,着重描述了从不同角度和层次进行来源数据管理的模式^[23]。

5.2 关联科学数据的管理和检索

因科学研究需要,在源数据项不变的情况下获取数据的方式可能会发生改变。J.Mwebaze等提出一种基于类别的对象版本管理框架,支持动态传递途径又能够进行独立性管理。该框架解决了数据流动过程中脚本的任意变化以及数据创建过程中各变动之间联系的管理问题^[24]。

S.Tuarob等介绍了ONEMercury。作为美国DataONE项目中获取全球环境和观测数据的工具,ONEMercury从多种数据存储库采集元数据并保证其可检索性。但是采集的元数据记录可能标注不清或缺乏有意义的关键词,很可能造成漏检。针对这种问题,他们开发了一套自动元数据注解算法,将问题转变为标签推荐问题,并给标签推荐提供一个分数传播模型。通过对四个环境科学元数据记录数据集进行实验,不仅验证了方法的有效性,也揭示了数据集的不同属性^[25]。

在复杂科学领域,如药理学,两个概念之间的操作等价(operational equivalence)经常是上下文特定、用户特定和任务特定的。现有的关联数据集成程序和等价服务并不考虑上下文和用户任务。因此,C.Brenninkmeijer等提出一种想法,通过在关联数据上运用“科学镜片”(scientific lenses),让用户控制操作等价概念。这种科学镜片会改变被激活的链接,这种链接发生在影响返回给用户数据的数据集之间^[26]。

N.Wilson等认为,为了更有效地收集世界生物多样性的观测结果,需要创建一个精确的、机器可理解的描述有机体群体形态的高效流程。由于形态描述需要以现代科学名称出版,这些描述在首次出版后再更改是很常见的,这就有可能导致数据丢失。针对这种情况,他们提出在提高观测数据采集质量的同时,采用语义白话系统(Semantic Vernacular System)来创建这类描述并管理它们与正式学名之间的关系^[27]。

5.3 科学信息的定量与定性集成

K.Page等提供了一个灵活的模式来核对和描述科学的语义环境。他们提出用一 workflow 描述一个实验,然而集合科学过程的数字化过程和产品,如输入输出数据、方法、软件、作用物、分析、传播、分享、再利用和这些集合资源的关联和关系等,则需要研究对象和一系列丰富的支持研究对象的工具来实现^[28]。

J.Zhao等人开发了一种轻量级命令行(command-line)工具——研究对象管理器(Research Object Manager),为科学家提供了一种聚合实验材料和研究方法的直接简单的方式,而这些实验材料和研究方法随后就可以出版并与同行分享,或关联到科学出版物,以增强实验结果的可重复性和可信性^[29]。

5.4 本体技术在科学信息上的应用

G.Palma等利用本体中的语义和共享注释语义从带注释的关联开放数据中进行模式挖掘,识别了基于共享概念和本体关联的一对概念中的注释信号。一个注释信号正式成为一个表示共享注释间关系边界的区分,在药品和基因数据集中进行的实验证明了用注释信号发现模式的有效性^[30]。

N.Adams等就如何以本体集合的方式公布一组数据和如何通过RESTful语义网服务来使用这些数据作

了报告。本体、实验数据和计算服务构成了语义可行的实验室笔记本所需的基本元素,促进多实验研究的同时减少其复杂性和错误率^[31]。

科学的两个重要特性是再现性和明确性,这两种特性与科学实践过程中良好的结构步骤和规范的描述密切相关。F.Osborne等提出一种作者引文分布聚类的新方法,从作者的角度研究不同资源的语义关联,并提出文献计量数据本体BiDO来对语义关系进行规范化表示^[32]。

5.5 数据语义化描述

随着关联开放数据的快速发展,大量公开数据中出现的语义描述缺乏问题越发凸显,虽然已有许多工具帮助用户快速将数据库转化为RDF,但并没有实现将数据简易映射到已有本体的方法。C.A.Knoblock等提出了一种允许用户交互式映射其结构化数据源到一个既有本体,然后再利用该映射生成RDF三元组的方法。这种方法自动生成从数据源到本体的映射,但由于准确完整的映射有时是模棱两可的,所以允许用户自行交互式定义其映射,随后以一个名为Karma的系统实现并验证了该方法的可行性^[33]。

T.Vision等探讨了语义描述集的相似性,强调了数据集背后的需求,包括速度、对大量数据集的可扩展性、显示统计和生物有效性、易用性等^[34]。O.Giraldo等提出了SMART协议,通过基于本体的方法对科学实验数据进行规范化语义表示^[35]。

5.6 科学信息的溯源、质量评估和共享研究

J.P.McCusker等认为,目前关联科学面临的一个挑战是如何在关联科学云中充分地描述一条信息的来源,而信息来源的表述是在科学应用中信任关联数据的关键。通过描述证明标记语言(Proof Markup Languages, PML)信息源利用概念,他们介绍了信息源模型并展示如何在书目资源的功能需求(Functional Requirements for Bibliographic Resources, FRBR)这一图书馆学模式中模型化^[36]。

起源模型对于描述实验结果至关重要。D.Garijo等提出用基于PROV扩展的P-PLAN模型来表示执行和描述执行自身源记录的计划。他们促进并讨论了以关联数据的形式使用P-PLAN和PROV发布科学工作流

程的方法,建议将P-PLAN作为一个公布计划的词表,将执行连接到P-PLAN上视为一个PROV标准的拓展。P-PLAN是通用的并可以拓展表示出更复杂的计划,它被认为是科学工作流公布的必要步骤,也是理解实验中方法执行的关键^[37]。

虽然研究过程中产生的数据已经能实现公开获取,但是通用资源发现机制的缺乏和格式化的互操作问题使得数据难以实现有效共享。Y.Gil等对科学数据共享进行研究,认为许多科学家不分享科学数据的原因是成本较高且传统的数据共享方法缺乏激励机制。他们在考虑科学文化实践的基础上提出了一种新的数据共享方法,该方法提供了一种语义框架,将数据集贡献与科学问题直接链接,通过让每个科学家贡献元数据降低了数据共享的负担,跟踪并公开所有贡献者的信誉。为了阐明其做法,他们描述了一种作为语义维基扩展的最初原型,不仅能够导入关联数据,而且能够以关联数据的形式发布任何用户创建的新内容^[38]。

J.Zhao等研究了语义网中科学数据信息的质量评估,提出一种基于清单的科学信息质量评估方法,允许特定质量要求的表达,帮助用户不受现有质量规模的束缚^[39]。

此外,还有学者对关联科学核心词汇和研究过程的可复制性进行了研究。B.Goncalves等提出,作为让科学资产显现出来并关联成数据网络的核心要素之一,“假设”这个词是关联科学核心词汇的一部分。他们通过追求最小的本体论承诺,阐述了假设的语义观点及其链接,将假设工程以关联数据的形式进行处理,并通过扩展该语义工程建立了关联科学核心词汇,旨在满足诸如计算科学等模型科学之词汇需要。然后,通过引出和关联来自于计算血液动力学领域已发表成果的假设,将扩展的关联科学核心词汇进行了例示^[40]。M.Hildebrand等介绍了利用语义网技术进行复制研究的实验过程。他们详细描述了对International Journal of Medical Sciences期刊上关于一项药物警戒研究的论文进行复制研究的具体步骤,其基础和关键是通过使用PROV-O本体来对计算工作流进行建模^[41]。

6 结语

本文基于关联科学国际研讨会论文对关联科学的发展现状进行分析,对比了4届会议的召开宗旨与目标、研讨话题和会议文献主题,梳理了关联科学的研究现

状与发展进程。从会议总体来看,国外关联科学研究重心从基础理论转移到了应用实证,从数据语义描述过渡到知识关联发现,并涉及地理学、生物医学、药理学、人文科学以及计算机科学等诸多领域。然而,关联科学的全面实现目前还存在缺乏统一机制、技术和资源基础建设不完全等诸多问题,这些问题的解决都依赖于相关方面的进一步研究。国内相关研究相对滞后,尚未形成规模,研究重心主要集中在理论探讨和技术实现上。国内学者应该以现有的研究和实践为基础,借鉴国外理论研究与实践成果,加快知识交流与共享的步伐。

参考文献

- [1] 张晓林. 颠覆数字图书馆的大趋势[J].中国图书馆学报, 2011, 37(5): 4-12.
- [2] 肖潇,吕俊生. E-science环境下国外图书馆科学数据服务研究进展[J].图书情报工作,2012,56(17): 53-58.
- [3] 唐义,张晓蒙,郑燃. 国际科学数据共享政策法规体系: Linked Science 制度基础[J]. 图书情报知识, 2013(3):67-73..
- [4] linkedscience [EB/OL]. [2015-07-20].<http://linkedscience.org/about/>.
- [5] Kauppinen T, Espindola G M. Linked open science-communicating, sharing and evaluating data, methods and results for executable papers [J]. Procedia Computer Science, 2011(4): 726-731.
- [6] 唐义,肖希明. 关联科学:一种全新的科研支撑方式[J].图书馆杂志, 2013, 32(8): 4-11.
- [7] linkedscience [EB/OL]. [2015-07-20].<http://linkedscience.org/events/lisc2011/>.
- [8] linkedscience [EB/OL]. [2015-07-20]. <http://linkedscience.org/events/lisc2012/>.
- [9] linkedscience [EB/OL]. [2015-07-20].<http://linkedscience.org/events/lisc2013/>.
- [10] linkedscience [EB/OL]. [2015-07-20].<http://linkedscience.org/events/lisc2014/>.
- [11] Groth P T, Gil Y. Linked Data for Network Science[C]//LISC. 2011.
- [12] Di Iorio A, Peroni S, Vitali F, et al. Semantic lenses to bring digital and semantic publishing together [J]. Linked Science, 12.
- [13] Más S, Müller M, Henzen C, et al. Linking the Outcomes of Scientific Research: Requirements from the Perspective of Geosciences[C]//LISC,2011.
- [14] Nguyen V, Bodenreider O, Minning T, et al. The Knowledge-driven Exploration of Integrated Biomedical Knowledge Sources Facilitates the Generation of New Hypotheses[C]//LISC, 2011.
- [15] Kazemzadeh L, Kamdar M R, Beyan O D, et al. LinkedPPI: enabling intuitive, integrative protein-protein interaction discovery[J]. Linked Science, 2014: 48.
- [16] Schneider J, Ciccarese P, Clark T, et al. Using the Micropublications ontology and the Open Annotation Data Model to represent evidence within a drug-drug interaction knowledge base[C]//Workshop on Linked Science 2014—Making Sense Out of Data at ISWC 2014,2014.
- [17] Jupp S, Malone J, Gray A J G. Capturing provenance for a linkset of convenience[J]. Linked Science, 71.
- [18] Nordhoff S, Hammarström H. Glottolog/Langdoc: Defining Dialects, Languages, and Language Families as Collections of Resources[C]//LISC,2011.
- [19] Meroño-Peñuela A, Ashkpour A, Rietveld L, et al. Linked humanities data: The next frontier? A case-study in historical census data[C]//Proceedings of the 2nd International Workshop on Linked Science 2012, 2012:951.
- [20] Sigarchian H G,Meester B D,Nies T D, et al. EPUB3 for Integrated and Customizable Representation of a Scientific Publication and its Associated Resources[C]//Proceedings of the 4th Workshop on Linked Science,2014.
- [21] Willemsen T, Feenstra A, Groth P. Building Executable Biological Pathway Models Automatically from BioPAX[C]// 3rd International Workshop on Linked Science 2013,2013:2.
- [22] McLean C, Gahegan M, Kubke F. Capturing intent and rationale for Linked Science: design patterns as a resource for linking laboratory experiments[C]// 3rd International Workshop on Linked Science 2013,2013:30.
- [23] Ockeloen N, Fokkens A, ter Braake S, et al. BiographyNet: Managing Provenance at multiple levels and from different perspectives [J].2013.
- [24] Mwebaze12 J, Boxhoorn D, Valentijn E. Supporting Scientific Collaboration Through Class-Based Object Versioning [J]. 2011.
- [25] Tuarob S, Pouchard L C, Noy N, et al. Onemercury: Towards automatic annotation of environmental science metadata[C]// Proceedings of the 2nd International Workshop on Linked Science, 2012.
- [26] Brenninkmeijer C, Evelo C, Goble C, et al. Scientific Lenses over Linked Data: An approach to support task specific views of the data. A vision[C]// Proceedings of 2nd International Workshop on Linked Science, 2012.
- [27] Wilson N, Wang H, McGuinness D. Scientific names and

- descriptions for organisms on the Semantic Web[C]//Proceedings of 2nd international workshop on linked science, 2012.
- [28] Page K, Palma R, Houbowicz P. From workflows to Research Objects: architecture for preserving the semantics of science[C]// Proceedings of the 2nd International Workshop on Linked Science, 2012.
- [29] Zhao J, Klyne G, Holubowicz P, et al. Ro-manager: A tool for creating and manipulating research objects to support reproducibility and reuse in sciences[C]//Proceedings of the 2nd International Workshop on Linked Science, 2012.
- [30] Palma G, Vidal M E, Raschid L, et al. Exploiting Semantics from Ontologies and Shared Annotations to Find Patterns in Annotated Linked Open Data[C]//3rd International Workshop on Linked Science 2013,2013:15.
- [31] Adams N, Haller A, Krumpholz A, et al. A semantic Lab Notebook— Report on a Use Case Modelling an Experiment of a Microwave-based Quarantine Method[C]//3rd International Workshop on Linked Science 2013, 2013:46.
- [32] Osborne F, Peroni S, Motta E. Clustering citation distributions for semantic categorization and citation prediction [J]. 2014.
- [33] Knoblock C A, Szekely P A, Ambite J L, et al. Interactively Mapping Data Sources into the Semantic Web[C]//LISC, 2011.
- [34] Vision T J, Blake J, Lapp H, et al. Similarity Between Semantic Description Sets: Addressing Needs Beyond Data Integration[C]// LISC, 2011.
- [35] Giraldo O, García A, Corcho O. SMART Protocols: Semantic Representation for Experimental Protocols [J]. Linked Science, 36.
- [36] McCusker J P, Lebo T, Ding L, et al. Where did you hear that? Information and the Sources They come From[C]//LISC, 2011.
- [37] Garijo D, Gil Y. Augmenting prove with plans in p-plan: scientific processes as linked data[C]. CEUR Workshop Proceedings, 2012.
- [38] Gil Y, Ratnakar V, Hanson P C. Organic data publishing: A novel approach to scientific data sharing[C]//Proceedings of the 2nd international workshop on linked science, 2012:951.
- [39] Zhao J, Klyne G, Gamble M, et al. A Checklist-Based Approach for Quality Assessment of Scientific Information[C]// 3rd International Workshop on Linked Science 2013, 2013.
- [40] Gonçalves B, Porto F, Moura A M. On the semantic engineering of scientific hypotheses as linked data[C]//Proceedings of the 2nd International Workshop on Linked Science, 2012.
- [41] Hildebrand M, Hoekstra R, Ossenbruggen J. Using Semantic Web Technologies to Reproduce a Pharmacovigilance Case Study[C]//3rd International Workshop on Linked Science 2013, 2013:72.

作者简介

刘磊, 男, 1954年生, 南京农业大学信息科学技术学院教授, 博士生导师, 研究方向: 数字资源建设, E-mail: llxx_2000@263.net。
郭诗云, 女, 1991年生, 南京农业大学信息科学技术学院硕士生, 研究方向: 信息组织, E-mail: guoshiyun91@163.com。

Study of Linked Science from the Perspective of International Workshop on Linked Science

LIU Lei, GUO ShiYun

(Department of Information Management, Nanjing Agricultural University, Nanjing 210095, China)

Abstract: With the development of linked data technology, scientific ways changed from digital research (e-science) to related sciences (Linked Science). By analyzing International Workshop on Linked Science, aimed at understanding the history and current development progress associated with Linked Science, and provide reference and inspiration for domestic. Through comparative analysis of the objectives and topics, the subjects were summarized as 6 topics including the associated application of science and case studies, quantitative and qualitative relevance of scientific data management and retrieval of scientific information integration, ontology in scientific information six aspects of applications, data, semantic descriptions and sources of scientific information, the quality and sharing. The results shows domestic research is still relatively weak, should further strengthen the research.

Keywords: Linked Science; Scientific Data; Semantic Web

(收稿日期: 2015-07-21; 编辑: 雷雪)