

基于知识组织体系的多维语义关联数据构建研究*

□ 鲜国建 赵瑞雪 孟宪学 朱亮 寇远涛 张洁 / 中国农业科学院农业信息研究所 北京 100081

摘要: 文章简要分析知识组织体系与关联数据的区别与联系,介绍了可在不同层次进行语义关联描述的本体模型和专业叙词表。面向科研创新信息需求,应用通用本体和农业科学叙词表等知识组织体系,建立了覆盖知识组织体系、科技文献和科学数据等多类资源的多维语义关联描述框架,实现了从资源的外部属性特征到内容层面的规范描述与语义关联。最后,以农业领域的信息资源为例,基于开源工具D2R,完成了关系型数据库中多类资源向多维语义关联数据的语义映射、自动转化与关联构建。

关键词: 关联数据, 知识组织体系, 本体, 叙词表, 科技文献, 科学数据

DOI: 10.3772/j.issn.1673—2286.2014.03.002

1 引言

作为在语义网中使用URI和RDF发布、分享、连接各类数据、信息和知识的最佳实践^[1],关联数据是实现海量、异源、异构信息精细化揭示、深度序化和语义化组织的有效途径,也是解决海量信息因离散孤立、缺乏语义而难以被计算机智能处理这一难题的有效手段。据统计,为便于计算机更加容易理解和处理数据,在地理信息、生命科学、大型传媒、商业企业、政府部门、学术出版和图书馆等领域已发布的近300个关联数据集中,近65%的关联数据集都采用了都柏林元数据DC、朋友的朋友FOAF、简单知识组织体系SKOS和地理本体GEO等被广泛使用的本体或通过词汇集来描述和关联数据^[2]。

深入分析已构建发布的关联数据集不难发现,大部分关联数据资源较为分散,资源聚合性较差,尤其是在面向科学研究提供关联数据服务时,尚未根据科研人员对科技信息的实际需求,从多维度进行专业领域多类型信息资源的整合和数据关联。此外,在建立数据间语义关联关系时,一般只重点考虑数据可关联的外部特征(如名称、时间、地点等),缺乏对数据内容尤其是专业领域的属性和语义关系进行描述和揭示,其中主要原因就是未充分利用专业叙词表等知识组织体系

和语义资源,导致语义关联层次不够深入。

本文接下来简要分析了知识组织体系与关联数据的区别与联系,介绍了可在不同层次进行语义关联描述的本体模型(通用词汇集)和专业叙词表。在此基础上,面向科研创新需求,基于通用本体和农业科学叙词表等知识组织体系,建立了覆盖知识组织体系、科技文献和科学数据等多类资源语义关联描述框架,并以农业领域科技信息资源为例,基于开源工具D2R,完成了关系型数据库中多类资源向多维语义关联数据的语义映射、自动转化与关联构建。

2 知识组织体系与关联数据

2.1 区别与联系

作为组织信息和知识的各类规范和方法的总称,知识组织系统(Knowledge Organization Systems, KOS)是获取、组织、管理和利用知识的重要手段^[3],既包括各种叙词表、主题词表、分类法等传统信息组织工具,也包括主题图和本体等新型知识组织技术。语义网络环境下,尤其是从传统文件网络向具有结构化和富含语义的数据网络(Web of Data)演进过程中^[4],知识组织系统又重新获得人们的广泛关注和高度重视,在组

* 本文是国家“十二五”科技支撑计划项目课题“基于STKOS的知识服务应用示范”(编号:2011BAH10B06)和公益性科研院所基本科研业务费课题的研究成果。

织管理、挖掘分析和开发利用海量信息资源实践中发挥日益重要作用。

关联数据为在网络上发布和链接结构化数据提供了新的载体和手段,使得机器能更准确理解和智能处理数据。然而,关联数据自身并不会给数据增添任何语义信息,而只是为更便捷地承载语义信息提供了基础框架。因此,富含语义关系的叙词表、本体等知识组织系统,与关联数据可有机融合、互为补充。一方面,可将知识组织系统转化、发布为关联数据,使其适应语义网络环境下新的发展和应用需求;另一方面,知识组织系统可在外部属性特征和专业领域实质内容等多个层次描述数据间的语义关系,为其提供明确规范的语义信息。二者相辅相成,可为构建富含语义的数据网络共同发挥作用。

2.2 通用本体与专业叙词表

由于关联数据不具备携带语义功能,因此,在构建和发布关联数据时,有必要应用被广泛使用的语义关联描述框架模型(本体或词汇集)和专业领域的叙词表,才能更规范、广泛、深入地实现数据间的关联和互操作。目前,研究社区已建立了不同层次的通用本体模型:

关联数据集词汇表(Vocabulary of Interlinked Datasets, VoID)和数据溯源词汇表(Provenance Vocabulary)等词汇集^[5],对关联数据生态系统中关联数据集的可用性、质量、性能和可靠性等问题进行了规范描述^[6],使得数据集的自动发现、筛选和查询优化等工作变得更加方便。

简单知识组织系统(Simple Knowledge Organization System, SKOS)、SKOS-XL和词汇表的朋友(Vocabulary of a Friend, VOAF)等本体框架,为叙词表、分类法、主题词表等知识组织体系提供一套规范、灵活、可扩展的描述转化机制,并提供了多种属性来定义词汇表之间的继承、扩展和关联等网络关系,以便实现各类知识组织系统资源的共享和重用。

都柏林核心元数据集(Dublin Core Metadata Initiative, DCMI)^[7]、书目本体(Bibliographic Ontology, BIBO)、FRBR书目本体(FRBR-aligned Bibliographic Ontology, FaBiO)和出版需求工业标准元数据(The Publishing Requirements for Industry Standard Metadata, PRISM)等为学术期刊、会议录、文集汇编等母体文献和篇级数据等都提供了通用、标

准的描述规范。

研究社区语义网本体(Semantic Web for Research Communities, SWRC)^[8]、科研本体VIVO Core和朋友的朋友本体FOAF等,为描述科研机构、科研项目、科研人员、科研成果和科研条件等与科研创新活动密切相关的各类核心对象实体和属性,以及数据间关联关系提供了规范的语义关联描述框架模型。

专业叙词表方面,以涉农领域为例,国内有中国农业科学院农业信息研究所研制的农业科学叙词表(Chinese Agricultural Thesaurus, CAT),收录了6万多个叙词概念并建立了13万余条词间语义关系。AGROVOC是由FAO更新维护的多语种叙词表,涵盖农业、林业、渔业、食品等领域,已收录3万多个概念,每个概念都用多达22种语言进行描述,并与国际上多个词表建立了语义关联,并基于RDF/SKOS-XL格式发布了关联数据版本^[9]。NALT是由美国国家农业图书馆等机构编制的农业叙词表,主要收录了农业、生物及相关领域的9万多个术语和4万多词间关系,提供了英语和西班牙语两个语种版本,该词表在2011年就将其发布为开放的关联数据。EUROVOC是由欧盟管理维护的多语种叙词表,最新版本中的语种多达22种,收录概念6883个,主要用于跨语言检索的词表。LCSH是美国国会图书馆以本馆的字典式目录为基础,以标题语言编制的美国国会图书馆主题词表。LCSH是世界上使用时间最长、范围最广、规模最大、影响最大的一部综合性标题表,2011年开始提供关联数据在线检索、浏览、解析和下载服务^[10]。

通过综合应用上述通用本体(词汇集)和专业叙词表,将能以更加通用、更加规范和更加科学的方式,在资源外部特征属性和实质内容主题揭示层面,实现知识组织体系与科技文献和科学数据等多类资源的多维语义关联描述和揭示,这也更加有利于提高关联数据的可见性和互操作性。

3 基于知识组织体系的多维语义关联描述机制

尽管不同来源、类型、结构、载体、用途和表现形式的信息资源可能存在这样或那样、显式或隐性的关联关系,但本文只重点开展与农业科研创新密切相关的科技文献、科学数据、农业领域知识组织体系等资源多维语义关联关系构建研究。其中,科技文献重点选择国家农业图书馆的农业科技学术期刊、图书、文集汇编

和会议录等类型，科学数据则包括国家农业科学数据共享平台中的科学数据库集及科技机构、科技人员和科研项目特色数据库^[11]。在关联和应用知识组织体系方面，作者在《农业科学叙词表关联数据构建研究与实践》一文中已实现了农业科学叙词表向关联数据的转化，并与AGROVOC和NALT等叙词表初步建立了语义映射和关联^[12]，本文将直接应用这一研究成果。

3.1 基于本体的科技文献外部属性关联

学术期刊、图书、会议录等科技文献为开展科研创新提供了重要信息支撑。尽管越来越多的图书馆已将其馆藏资源以关联数据的方式发布，如瑞典联合目录LIBRIS、美国国会图书馆发布SKOS版本的LCSH，以及OCLC基于WorldCat.org发布的书目关联数据等，但主要都集中在书目层面，而对篇级文摘和引文数据等更具价值的信息在更小粒度和内容层面的揭示关联还有待深入推进。在所有科技文献中，学术期刊是最为复杂的一类，因为涉及期刊品种、卷期、摘要、作者、引文等多类对象信息。本文综合应用DCMI、BIBO、PRISM、SWRC及FaBIO等被广泛使用的本体（词汇集），对国家农业图书馆的学术期刊等科技文献的外部特征进行了规范化表达和语义描述，其中：

期刊母体类可用bibo:Journal、swrc:Journal和fabio:Journal进行规范描述。该类的属性主要包括期刊名称、ISSN、语种、起始年份、出版商等对象属性和数据属性，其规范化描述如图1所示。

如图1所示，大部分期刊母体元数据项都可直接复用现有本体来表示，且有多个本体模型可供选择使用。然而，作者并未找到能准确表达期刊母体起始出版年份、结束（停刊）年份的词汇。因此，在前缀agribibo的限定下，作者自定义了startPublishYear和endPublishYear两个词汇。此外，还自定义了对象属性hasIssue来显性地揭示期刊母体与其所有卷期的关联关系。严格意义上，出版商和出版地也可定义为对象属性，其属性取值则应当是一个对象类（如foaf:Organization和event:place）的实例，本文暂未对这两类信息单独描述。

单本期刊（卷期）类可用bibo:Issue和fabio:JournalIssue进行规范描述。该类的属性主要包括出版年份（swrc:year、fabio:hasPublicationYear）、卷（bibo:volume、swrc:volume）、期（bibo:issue、prism:number、swrc:issue）等属性，本文自定义了agribibo:isIssueOf对象属性，与期刊母体对象属性agribibo:hasIssue为互逆属性。一般只需在两个类中定义其中一个属性，就可通过推理得出另一属性的关联信息。

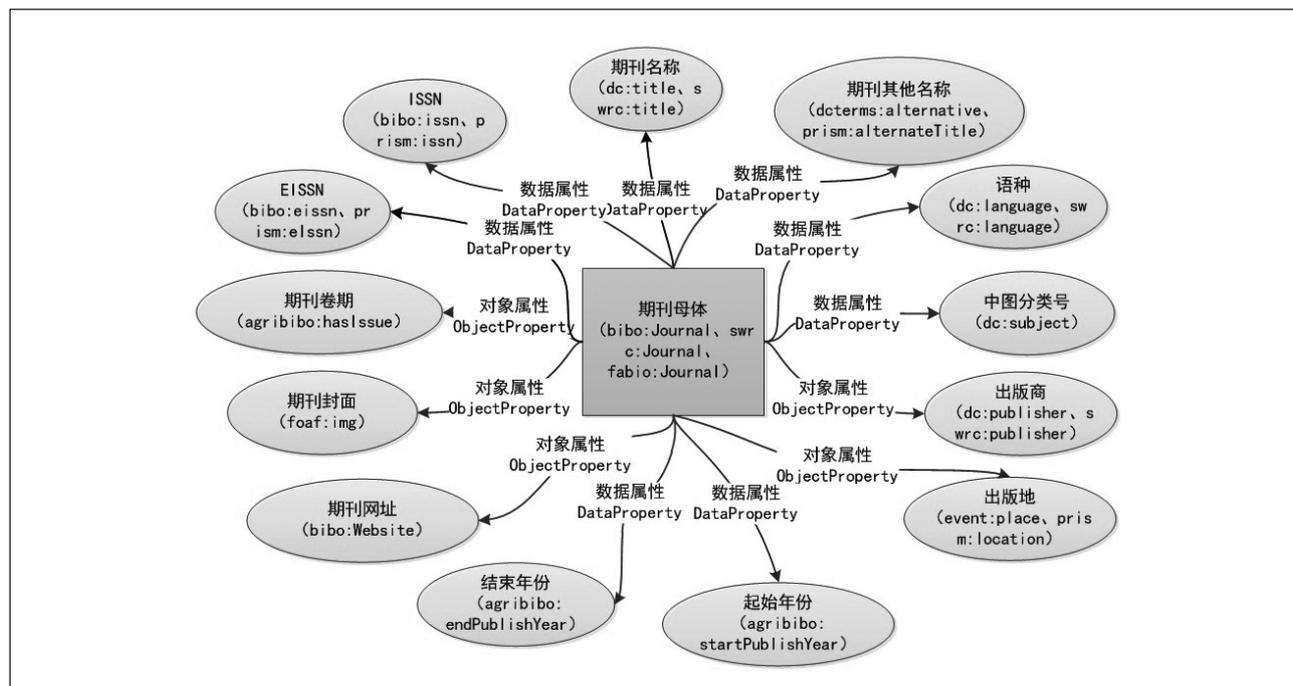


图1 科技期刊母体类及其核心属性

表1 学术期刊文摘类核心属性

序号	核心属性	属性类型	可选用描述词汇
1	所属期刊卷期	ObjectProperty	dcterms:isPartOf、agribibo:isJournalArticleOf
2	题名	DataProperty	dc:title、swrc:title
3	其他题名	DataProperty	dcterms:alternative、prism:alternateTitle
4	作者	ObjectProperty	dc:creator、foaf:maker、swrc:creator
5	关键词	ObjectProperty/DataProperty	dc:subject、swrc:keywords、prism:keyword
6	摘要	DataProperty	bibo:abstract、dcterms:abstract、swrc:abstract
7	语种	DataProperty	dc:language、swrc:language
8	起始页码	DataProperty	bibo:pageStart、prism:startingPage
9	结束页码	DataProperty	bibo:pageEnd、prism:endingPage
10	总页数	DataProperty	bibo:numPages、prism:pageCount
11	DOI	ObjectProperty	bibo:doi、prism:doi
12	SICI	DataProperty	bibo:sici、fabio:hasSICI

期刊文摘类可用**bibo:AcademicArticle**、**fabio:JournalArticle**和**swrc:Article**进行规范描述。该类属性主要包括所属期刊卷期、题名、作者、摘要等多个属性，其规范化描述如表1所示。

其中，文摘元数据中包括期刊卷期的唯一标识符ID，根据关联数据构建原则，本文将所属期刊卷期描述为对象属性，将通过一定规则直接与期刊卷期类的实例进行关联。事实上，对于作者信息，本文将作者姓名及其单位等信息抽象为作者类，通过作者对象属性在文摘类与作者类间建立关联。如果一篇论文有多个作者，则这种关联关系会出现多次。关键词属性既可处理为数据属性（**swrc:keywords**、**prism:keyword**），也可描述为对象属性（**dc:subject**）。后者需对关键词基于叙词表作进一步标引，将其标引为叙词表中的概念（见3.3节）。

作者类可用**foaf:Person**和**swrc:AcademicStaff**等进行规范描述。对该类初步设计了姓名（**foaf:name**、**swrc:name**）、所属机构（**swrc:address/swrc:affiliation**、**foaf:Organization**）和电子邮箱（**swrc:email**、**foaf:mbox**）等属性。目前，从学术论文中能提炼出有关作者的信息非常有限，主要包括姓名、所属机构（通讯地址）和电子邮箱等。由于之前对有关作者详细信息的需求并不多，以及加工成本等因素，已加工的数据中基本上未对作者姓名、所属机构、通讯地址和邮编等信息进行细分和规范加工，只是简单地堆放在一起。

在关联数据的环境下，更合理的方式是将作者类的属性进一步规范和扩展。

与学术期刊相比，其他文献类型的规范描述就相对简单，图书、文集汇编和会议录等对象类及部分特殊属性的规范描述如表2所示。

3.2 基于本体的科学数据外部属性关联

科学数据是指人类社会从事科技活动所产生的基本数据，以及按照不同需求而系统加工整理的数据产品和相关信息^[13]。科学数据资源是国家科技创新和发展需求的一种战略资源，是科技进步与创新的强有力支撑。当前，以数据密集型计算为特征的科学研究“第四范式”正在兴起^[14]，海量科学数据资源的开发利用也备受关注和高度重视。农业科学数据共享中心作为国家科技基础条件平台建设项目，目前共整合了作物科学、动物科学等12大类农业核心学科的数据库（集）700多个，数据量近3TB，形成了较为系统全面、有较高科学价值的科学数据资源库。

农业科学数据共享中心的数据库集本质上是相对独立的数据库。本文通过继承复用现有的DCMI、VIVO和FaBIO等本体模型，将数据库集类可用**vivo:DataSet**和**fabio:Database**进行规范描述。该类包括数据库中、英文名称、摘要、语种、关键词、负责单位、负责人等13个核心对象属性和数据属性，其规范化描述如图2所示。

表2 常见文献类型对象类及属性规范描述

科技文献类型	对象类可用本体词汇	可选用描述词汇
图书对象类	bibo:Book、swrc:Book、fabio:Book	ISBN(bibo:isbn、prism:isbn)、责任者 (bibo:editor、swrc:editor)、版次 (prism:versionIdentifier)
文集汇编类	bibo:Collection和swrc:Collection	
会议录类	bibo:Proceedings、swrc:Proceedings和fabio:ConferenceProceedings	所属学术会议 (event:produced_in)
学术会议录	bibo:Conference和swrc:Conference	会议名称 (swrc:eventTitle)、主办单位 (prism:organization、foaf:Organization)
文集汇编文摘类	bibo:Article和swrc:InCollection	
会议录文摘类	bibo:Article、swrc:InProceedings和fabio:ConferencePaper	所属会议录 (bibo:reproducedIn)、所属学术会议 (bibo:presentedAt)

图2中,对农业科学数据库集的核心元数据项进行了规范化描述。其中,将数据库集负责单位与负责人通过对象属性 (dc:contributor) 分别与科研机构 and 科研人员类建立了关联关系,而不仅是之前简单的字符串描述。针对数据集的关键词属性,本文同样采用了两种方式,既可处理为数据属性 (swrc:keywords、prism:keyword),也可对关键词基于农业科学叙词表作进一步标引,并通过对象属性 (dc:subject) 与标引结果建立关联。

与虚拟国际规范文档VIAF^[15]类似,在农业科学数据库(集)中,收集了与科技创新密切相关的农业

科技机构、科技人才和科研项目等规范数据,这些数据之间以及与科技文献之间都存在非常紧密的关联关系。其中,农业科技机构类可用swrc:Organization和vivo:Organization进行规范表达,农业科技人才类可用swrc:FacultyMember、vivo:FacultyMember和foaf:Person等表达,农业科研项目类可用swrc:ResearchProject和vivo:Project等类表达。限于篇幅,更详细的属性描述和关联不再详述。

3.3 基于叙词表的专业内容标引与关联

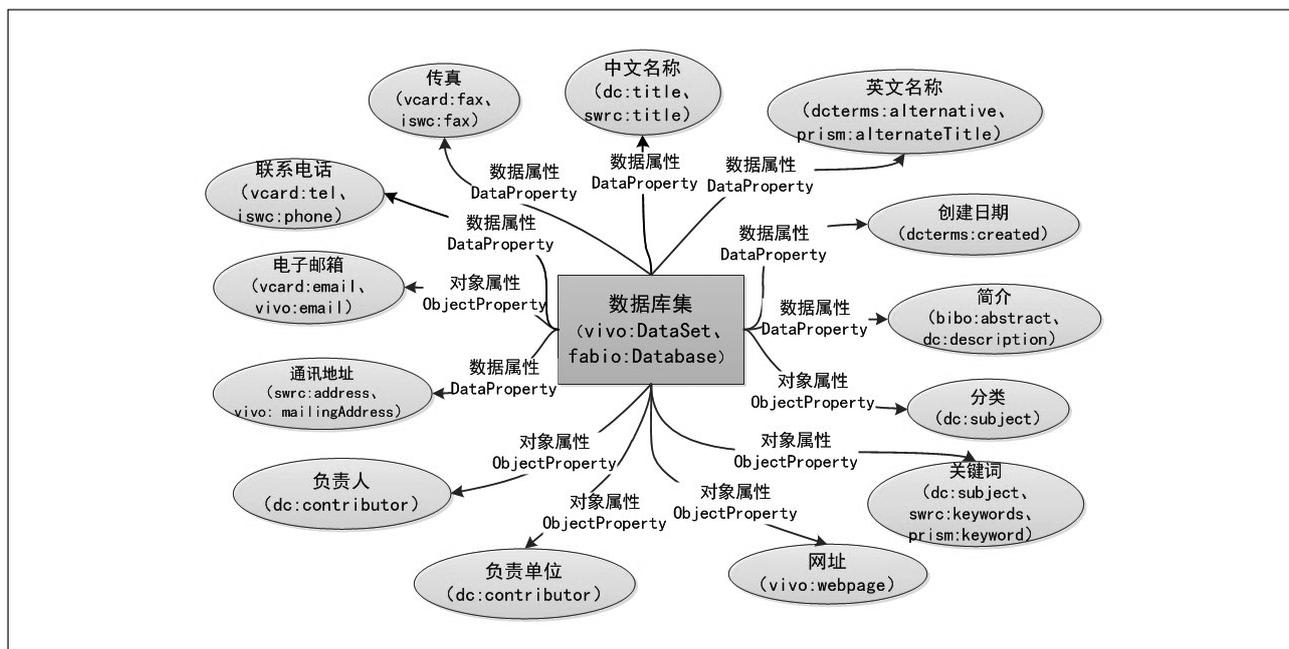


图2 数据库集类及其核心属性

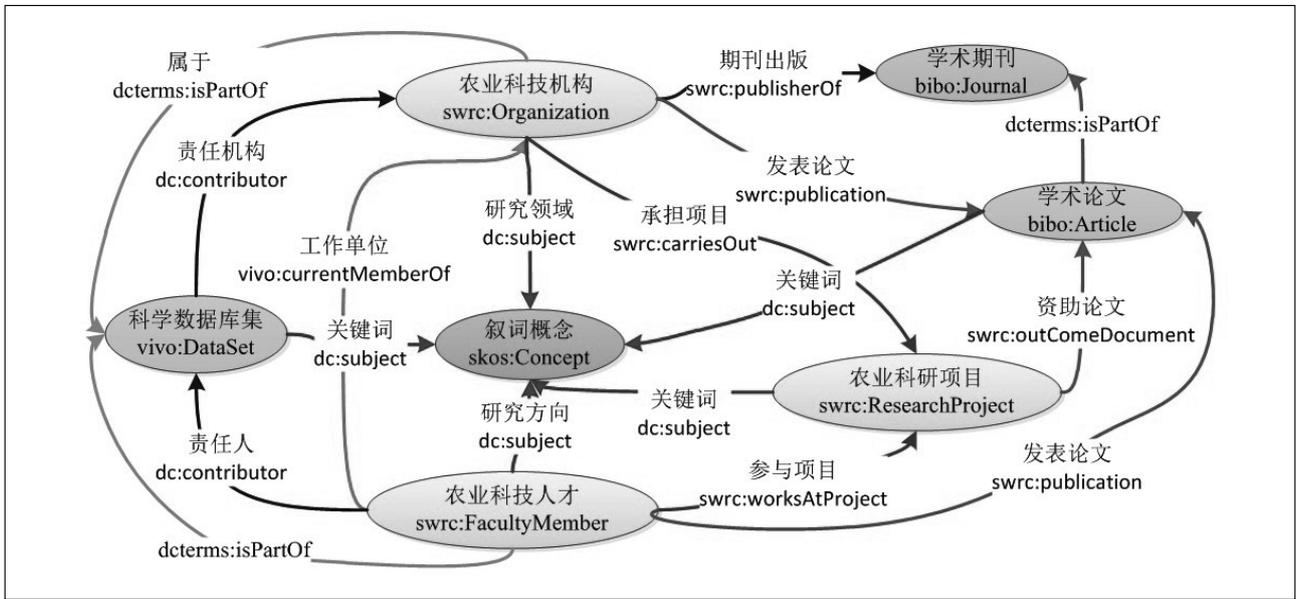


图3 科学数据与科技文献的语义关联模型

尽管基于上述通用本体或词表集可完成对科技文献和科学数据外部属性的规范描述和语义关联,但不同类型数据资源之间,尤其是在内容层面仍缺乏明确、规范和显性的关联揭示。为充分利用农业科技叙词表从内容层面丰富各类资源之间的语义关联关系,基于农业科学叙词表中丰富的概念和语义关系,开发了自动标引工具,初步实现各类资源在概念层面的自动标引(关联规则主要是基于关键词和叙词概念的精确匹配),也初步实现了叙词表、科学数据和科技文献之间的多维语义关联,为开展更丰富的知识服务应用奠定了语义基础。

3.4 多类型资源多维语义关联描述框架

本文在学术期刊、数据库集、科技机构、科技人员、科研项目等多类型资源间建立关联关系的基础上,还与农业科学叙词表中的规范概念建立了语义关联,这为基于农业科学叙词表中丰富的语义关系,更好地整合和挖掘资源奠定了语义基础。这也是本文在构建关联数据过程中采用的具有创新性的方法。图3展示了各类数据资源的多维语义关联关系。

其中,通过dc:contributor属性建立了科学数据库集(vivo:DataSet)与其责任单位(科技机构类,swrc:Organization)和责任人(科技人员类,swrc:FacultyMember)之间的关联关系;同时,dcterms:

isPartOf建立了科技机构及科技人员与它们所属的数据库集建立了关联。vivo:currentMemberOf属性建立科技人员与其所属科技机构的关联;swrc:carriesOut和swrc:worksAtProject属性分别关联了科技机构和科技人员参与的科研项目;swrc:publication属性揭示了科技机构和科技人员发表的学术论文(还可以关联更多科技文献);swrc:outComeDocument属性建立了科研项目与受其资助发表的学术论文的关联关系。

4 基于D2R的多维语义关联数据构建与发布

4.1 关联数据构建与发布关键流程

在构建与发布关联数据过程中,除构建上述语义关联模型外,最为关键的步骤还包括:

- 实体命名:即为每个实体建立稳定、可访问、可解析的唯一标识符HTTP URI生成机制。欧盟的欧洲行政互操作项目ISA发布的关于URI的调查报告,总结了在设计永久唯一标识符的10条原则^[16],其中第一条原则就是尽可能早地为资源分配唯一标识符。Dodds L等研究人员也总结了等级型(Hierarchical URIs)、继承型(Natural Keys)和重构型(Rebased URI)等8种常见的URI生成方式^[17]。

- 实体RDF化:采用RDF对每个对象实体及其属

性进行规范化、结构化的语义描述,使得对实例的描述能被计算机理解;

- 实体关联化: 采用RDF链接来描述各类实体对象之间的关联,并尽可能多地与外部数据源建立丰富的关联关系,使数据集具有跨实体发现的能力;
- 实体发布: 部署关联数据发布服务器,对外提供关联数据服务,根据内容协商机制返回正确的网页描述和RDF描述;配置SPARQL服务端(SPARQL endpoint),对外开放SPARQL 语义查询接口,供远程调用本地数据。

4.2 基于D2R的关联数据构建与发布

在前面设计和建立多维语义关联描述框架的基础上,接下来将遵循Tim Berners-Lee提出的创建关联数据应坚持的四项基本原则^[18],基于开源工具D2R,将存放于关系型数据库中的各类资源以关联数据的形式进行动态关联与发布。D2R的主要功能是以RDF视图和方式,实现对关系型数据库中的数据进行查询访问,并为RDF专用浏览器、SPARQL查询端以及HTML传统浏览器提供数据调用接口。作者基于D2R提供的D2RQ映射语言,完成了存放于数据库中的叙词表、科技文献和科学数据由表、列、表间关系向对象类、核心

属性及其关联关系的语义映射,编制了语义映射文件mapping-agridatas.ttl,图4是学术期刊、卷期、论文和作者几类对象的映射框架:

基于D2R Server,初步构建的农业科技文献语义关联数据构建与发布平台如图5所示。



图5 多维语义关联数据构建与发布平台

5 结语

本文基于本体和叙词表等知识组织体系,初步从多种科技文献和科学数据的外部特征属性到实质内容进行了规范描述和语义关联。将图书馆各类科技文献资源进行规范化描述和语义化关联组织,并将特色馆藏

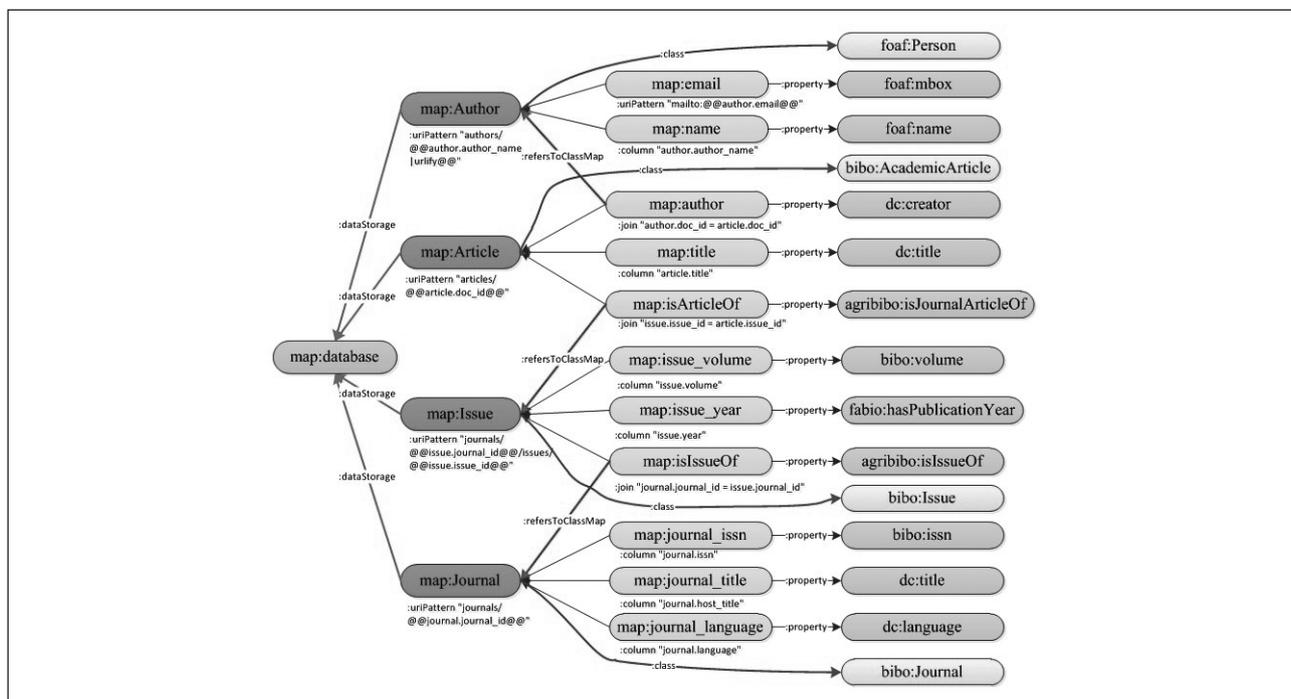


图4 期刊文献类及属性映射框架

资源发布成关联数据,可增加用户返回图书馆的途径,显著提高馆藏资源的可知性、可见度和可获得性。同时,基于知识组织体系将科学数据与科技文献关联起来,能有效整合、盘活、挖掘和利用好极具科学价值的科学数据资源,有效支撑数据密集型科研创新,最大限度地发挥科学数据的价值。然而,本文在语义关联的广度和深度还远远不够,知识组织体系的作用也还有待进一步挖掘,相关工作将在后续研究深入开展。

参考文献

- [1] Bizer C, Heath T, Berners-Lee T. Linked Data - The Story So Far[J]. International Journal on Semantic Web and Information Systems, 2009, 5(3): 1-22.
- [2] BIZER C, JENTZSCH A, CYGANIAK R. State of the LOD Cloud [EB/OL]. (2011-09-19) [2014-02-19]. <http://lod-cloud.net/state/>.
- [3] 王军,张丽.网络知识组织系统的研究现状和发展趋势[J].中国图书馆学报,2008,34(1):65-69.
- [4] LEE T B. Linked Data - Design Issues [EB/OL]. (2006-07-27) [2014-02-11]. <http://www.w3.org/DesignIssues/LinkedData.html>.
- [5] ALEXANDER K, CYGANIAK R, HAUSENBLAS M, et al. Describing Linked Datasets -- On the Design and Usage of void, the "Vocabulary Of Interlinked Datasets" [C]// Proceedings of LDOW2009, April 20, 2009, Madrid, Spain.
- [6] Golbeck J. Weaving a web of trust [J]. Science, 2008, 321(5896): 1640-1641.
- [7] DCMI Metadata Terms [EB/OL]. (2012-06-14) [2014-02-13]. <http://dublincore.org/documents/2012/06/14/demi-terms/>.
- [8] SURE Y, BLOEHDORN S, HAASE P, et al. The SWRC ontology - Semantic Web for research communities [C]// Proceedings of the 12th Portuguese Conference on Artificial Intelligence (EPIA 2005). Springer, Covilha, Portugal, December 2005.
- [9] AGROVOC Linked Open Data [EB/OL]. (2013-03-20) [2013-04-09]. <http://aims.fao.org/standards/agrovoc/linked-open-data>.
- [10] Library of Congress Subject Headings [EB/OL]. (2011-04-26) [2014-02-20]. <http://id.loc.gov/authorities/subjects.html>.
- [11] 国家农业科学数据共享中心[EB/OL]. (2011-12-02) [2014-02-17]. <http://www.agridata.cn/>.
- [12] 鲜国建,赵瑞雪,等.农业科学叙词表关联数据构建研究与实践[J].现代图书情报技术,2013(11):8-14.
- [13] 黄鼎成,郭增艳.科学数据共享管理研究[M].北京:中国科学技术出版社,2002.
- [14] HEY T, TANSLER S, TOLLE K. The Fourth Paradigm: Data Intensive Scientific Discovery [M]. Microsoft Research Publishing, 2009.
- [15] VIAF: The Virtual International Authority File [EB/OL]. (2010-06-15) [2012-02-21]. <http://viaf.org/>.
- [16] ARCHER P, GOEDERTIER S, LOUTAS N. Study on persistent URIs, with identification of best practices and recommendations on the topic for the MSs and the EC [ER/OL]. (2012-12-17) [2014-02-16]. https://joinup.ec.europa.eu/sites/default/files/D7.1.3%20-%20Study%20on%20persistent%20URIs_4.pdf.
- [17] DODDS L, DAVIS I. Linked Data Patterns: A pattern catalogue for modelling, publishing, and consuming Linked Data [EB/OL]. (2012-05-31) [2014-02-19]. <http://patterns.dataincubator.org/book/linkeddata-patterns.pdf>.
- [18] BIZER C, HEATH T, LEE T B. Linked Data - The Story So Far [M]. HEATH T, HEPP M, BIZER C. Special Issue on Linked Data, International Journal on Semantic Web and Information Systems (IJSWIS 2009).

作者简介

鲜国建(1982-),男,博士。研究方向:知识组织、关联数据、数字资源加工、信息系统开发,发表学术论文10余篇。E-mail: xgj@mail.caas.net.cn

Construction of Multidimensional Semantic Linked Data Based on Knowledge Organization System

Xian Guojian, Zhao Ruixue, Meng Xianxue, Zhu Liang, Kou Yuantao, Zhang Jie / Agricultural Information Institution of CAAS, Beijing, 100081

Abstract: This paper analyzes the difference and relation between knowledge organization system (KOS) and linked data, and also gives brief introduction about some well-known ontologies and thesauri. A multidimensional semantic linking framework covering KOS, literature and scientific data is designed based on some widely used ontologies and thesauri, to formally describing and semantically linking several types of resources both from outside properties and inside contents. Finally, some agricultural sample data in relational database is semantically mapped and automatically converted and linked as linked data based on the open source tool D2R.

Keywords: Linked data, Knowledge organization system, Ontology, Thesaurus, SciTech literature, Scientific data

(收稿日期: 2014-02-20)