

# 面向农业产业知识服务的科学数据与 科技文献关联研究与实践

柴苗岭<sup>1,2,3</sup>, 邹弈星<sup>4</sup>, 谭荣志<sup>3\*</sup>, 曾 怡<sup>5</sup>, 任运月<sup>6</sup>

(1. 中国科学院 成都文献情报中心, 成都 610041; 2. 中国科学院大学 经济与管理学院, 北京 100190;  
3. 中国科学院 山地灾害与地表过程重点实验室, 成都 610041; 4. 四川省农村科技发展中心, 成都 610042;  
5. 香港城市大学 商学院, 香港 999077; 6. 四川大学 公共管理学院, 成都 610065)

**摘要:** [目的 / 意义]为解决当前农业科技管理中广泛存在数据链不完整, 数据支撑决策能力不足的情况, 本文提出了一套从农业科技管理需求出发的产业知识服务模型, 并就其中的关键环节科学数据和科技文献关联进行了研究和实践, 以为图书情报机构服务农业科技管理和产业发展提供理论与实践案例。[方法 / 过程]本文首先梳理了知识服务和数据关联现状, 其次在分析农业科技管理中的产业需求特征和数据管理生命周期的基础上, 构建了包含产业特征的知识服务模型, 然后以四川农业科学数据和科技文献为基础, 从产业要素出发, 设计了本体模型和系统, 最后完成实践。[结果 / 结论]研究最后集成 20 个数据类型, 24 284 条科学数据和科技文献, 1 个系统平台和 1 个子平台, 建立了多源异构、中英两语种的数据关联, 实现了基于产业知识服务的科学数据和科技文献关联和平台交流。研究成果将对农业科技管理与图书情报学跨领域合作, 科学数据和科技文献跨部门共享提供借鉴。

**关键词:** 农业产业; 科学数据; 科技文献; 数据关联; 本体; 产业链; 大数据

**中图分类号:** G353.1      **文献标识码:** A      **文章编号:** 1002-1248 (2022) 03-0037-14

**引用本文:** 柴苗岭, 邹弈星, 谭荣志, 等. 面向农业产业知识服务的科学数据与科技文献关联研究与实践[J]. 农业图书情报学报, 2022, 34(3): 37-50.

**收稿日期:** 2021-07-23

**基金项目:** 中国科学院“山地灾害与地表过程重点实验室开放基金”; 中国科学院文献情报能力建设专项“开放知识资源体系建设(二期)”(Y1755); 四川省科技计划国际合作项目“面向知识服务的农业科学数据和科技文献关联融汇服务系统研究与示范”(2017HH0094)

**作者简介:** 柴苗岭, 博士研究生, 副研究馆员, 中国科学院成都文献情报中心, 研究方向为资源建设、知识组织、知识产权情报研究。邹弈星, 硕士, 副研究员, 四川省农村科技发展中心, 研究方向为农业科技管理。曾怡, 博士研究生, 香港城市大学商学院信息系统系, 研究方向为知识组织技术、知识管理。任运月, 博士研究生, 四川大学公共管理学院图书馆学, 研究方向为信息管理方法与应用

\*通信作者: 谭荣志, 工程师, 中国科学院水利部成都山地灾害与环境研究所, 研究方向为数据资源管理、山地灾害防灾减灾。Email: tanrz@imde.ac.cn

## 1 引言

随着大数据相关理论、技术和机制的逐步发展,各行各业越来越重视在管理决策中引入多源异构数据进行融合分析,进而提供大数据驱动下的决策咨询。从科技管理实践、科技管理与图书情报学科交叉角度来看,实现这一目标还需要在知识服务能力上下足功夫。其中产业服务是科技管理的重要职能之一,全产业要素下的数据挖掘及利用,有利于制定产业技术创新战略、规划。而科学数据和科技文献关联研究能够挖掘科学数据之间、科学数据和科技文献之间的深层次关系,有助于产业全域数据的整合和挖掘。

农业科技管理部门看到了这些问题,也建立了科学数据平台支撑管理。以四川省农业科学数据建设情况为例,作为农业大省和科技资源大省,四川省相关政府机构、科研机构支持建设了大量的数据和资源平台,为农业科技管理提供了良好的基础,如“四川省科技成果转化信息服务平台”“四川育种攻关数据共享平台”“农业病虫害多媒体数据库”等。但仍然存在数据量不足、数据孤岛、数据挖掘不够的现象。具体表现在3个方面:一是知识服务导向不明确;二是数据共享性较差;三是缺少科技文献。

面向上述问题,本文以现有科技管理中的产业服务特征为出发点,提出了基于农业产业服务的知识服务模式,就“产业服务”和“数据关联”两个关键问题构建了模型和平台,并用四川省农业科学数据和科技文献为核心数据集进行了建设实践。本研究为农业产业知识服务中的科学数据和科技文献关联提供了案例,为进一步开展全产业链下的农业知识服务做了基础性研究。

## 2 研究现状

### 2.1 农业知识服务研究现状

知识服务最早由任俊为引入图书情报行业<sup>[1]</sup>,张晓林进一步揭示了知识服务概念的本质,认为知识服务

要有效支持知识应用和知识创新的服务<sup>[2]</sup>。随着研究的深入,李霞等<sup>[3]</sup>提出将知识服务的模式分为专职顾问服务模式、参考咨询服务模式和自助服务模式3种。此外,柯平<sup>[4]</sup>将图书馆转型发展的行业环境划分为信息服务时代、前知识服务时代和后知识服务时代。行业环境的变化对数据价值挖掘和知识服务提出更高的要求。农业科学数据的价值在农业科技管理过程中不断凸显,涌现出各类农业知识服务模式。刘健宏等<sup>[5]</sup>将大数据背景下农业分学科知识服务模式归纳为“学科馆员+团队首席+研究团队”模式、“微信+信息平台”模式和“知识社区+专题信息”3类。王丹阳<sup>[6]</sup>从农业知识服务平台建设实践出发,将知识服务划分为政策导向、产品导向、技术导向、市场导向和流量导向。

### 2.2 科学数据现状与科技文献关联方法研究现状

#### 2.2.1 科学数据研究现状

科学数据的研究最早可追溯到1960年美国成立国家大气研究中心,开展地球科学数据的建模、收藏和保存工作<sup>[7]</sup>。国外的农业领域,美国建立了较为全面的农业科学数据资源系统和农业计算机网络系统,欧盟、德国等国家亦致力于科学数据开放共享<sup>[8]</sup>。研究内容上,侧重元数据体系结构的建立<sup>[9]</sup>、基于物联网开发数据平台<sup>[10]</sup>、利用商业智能集成农业数据<sup>[11]</sup>、开放信息系统模型框架和农业资源平台关键的硬件、软件技术<sup>[12]</sup>研究等。此外,数据管理层面的数据管理人员及合作<sup>[13]</sup>研究、数据互操作<sup>[14]</sup>、跨学科数据管理<sup>[15]</sup>及数据开放安全性和隐私问题<sup>[15]</sup>也受到国外学者关注。

国内的科学数据研究可追溯到2001年国家科学数据共享工程启动气象科学数据试点,它先后带动了资源环境、农业、人口与健康、基础与前沿等领域科学数据建设与共享<sup>[16]</sup>。国内研究总体而言涵盖了资源整合<sup>[17]</sup>、数据存储<sup>[18]</sup>、数据监管<sup>[19]</sup>、数据汇交与管理<sup>[20]</sup>、数据开放与出版<sup>[21]</sup>、数据重用<sup>[22]</sup>等环节,核心是农业数据资源整合和使用。

#### 2.2.2 科学数据与科技文献关联方法现状

《科学数据管理办法》将科学数据划分为科学研究

活动、基础研究、应用研究和试验活动的数据<sup>[23]</sup>。研究采用该分类,对科学数据的定义、产出和保存机构、数据载体进行调研和分析,认为相关机构主要有科研院所和大学(包括中心、实验室、观测站等),图书情报机构、政府管理部门和企业,科学数据载体主要有数据集、数据库、期刊、专著、专利、报告、政策法规等。需要指出的是,从载体的角度来看,科技文献也是科学数据的一种。

数据关联方面,TONY等<sup>[24]</sup>提出了数据密集型科学发现,构建了科学数据和科技文献关联的使用场景,认为科学数据关联能够提高科学的“信息速率”,促进科研人员的科学生产力。本文从关联技术角度进行分析,将关联方法划分为以下4类。

(1) 元数据关联。是基于科学数据和科技文献数据的外部 and 内部特征描述,寻找两类数据的关关节点,形成物理关联。如黄筱瑾<sup>[25]</sup>从元数据描述元素出发,总结了科学数据和科技文献的作者关联、学科分类号关联、关键词关联模式。

(2) 语义关联。是从语义层面发现实体、找出关联关系,增强数据共性的方法。如鲜国建<sup>[26]</sup>从科学数据语义关联的方法和技术出发构建了农业科技多维语义关联模型和关联数据驱动的知识服务系统。

(3) 算法关联。是通过数学公式,判断数据间的相关关系及其距离的方法。如李帆<sup>[27]</sup>选取FP-Growth算法进行属性间趋势关联性挖掘,从列车监测数据出发通过部件间的状态关联性发现其他部件可能存在的问题。

(4) 引用关联。引用关联即利用科技文献和科学数据的引用关系,构建引文网络<sup>[28]</sup>。是一种直接、基本的数据关联方式。如丁文姚等<sup>[29]</sup>结合国家标准《信息技术科学数据引用》的引用元素并应用统计学方法描述其科学数据引用特征,揭示了特征间的关联关系。

农业科学数据作为战略性、基础性的科技资源,其有效研究和利用将有助于农业产业视角下的全数据融合<sup>[30]</sup>。农业产业知识服务需求贯穿于全产业链服务,可以判断出服务主体对科学数据的需求将涵盖多源、多类型、非结构化的数据。本体作为基于语义的知识组织方法已经被广泛地研究与应用,本研究在产业管

理特征上采用本体方法关联语义有一定的研究意义和复用价值。

## 3 研究设计

### 3.1 数据遴选及特征分析

数据关联关键环节之一是对数据源的遴选及数据分析,寻找关联点。

研究以四川农业科学数据为基础,构建核心数据集,遴选导向是科学数据和文献数据、数值数据和文本数据、中文数据和英文数据。研究数据选择了农业科研管理机构、图书情报机构、科学研究机构/科学数据共享平台的数据,遴选了16类科技管理中常用数据和4类科技文献数据,然后对这20种数据的元数据进行分析,发现关联特征(表1、表2)。

从表1和表2可见,元数据内可以概括为描述客观存在的外部特征和表达内容的内部特征。由于数据类型多、结构不同,通用的外部和内部特征较少,数据之间的相似性不强,如果要进行数据关联,需要补充科学数据和科技文献的共性特征,如数据的领域分类、产业环节分布、地理位置描述、语种描述等。

### 3.2 基于科技管理的知识服务特征及服务需求分析

#### 3.2.1 知识服务特征

(1) 数据使用受国家政策指导。管理部门在数据的使用上,侧重统计数据、成果转化、人才数据的使用,具有产业化管理特征。这一特征由管理部门的性质决定,具体表现是使用宏观、中观数据的带有统计意义的数据,重视成果转化及人才,并且对科学研究内容、细节等不做追求。相关数据类型有政策法规数据、科技项目数据、获奖成果数据、专家数据。在近年数据使用上,除了支撑产业管理,也体现了科技扶贫工作特色,具体表现是对科技专家、三区人才、科技特派员的数据采集。

(2) 数据加工关注产业链环节。在农业产业管理

表 1 农业科学数据元数据结构表

Table 1 Metadata schema of agricultural science data

| 资源类型                     | 元数据   | 数据来源           |
|--------------------------|---|----------------|
| 科技计划项目                   | 名称、负责人、机构、类型、立项年度、地点  | 四川省农村科         |
| 新品种新产品                   | 品种名称、证书编号、所有者、品种来源、品种特性、栽培要点、特征特性、内在品质、产量表现、审定年份、适宜范围、地点                            | 技发展中心          |
| 科技专家/科技特派员               | 姓名、职位职称、所在地区、工作单位、从事专业  |                |
| 三区人才                     | 姓名、性别、专业技术职称、行业、专业领域、工作单位、派往市州  |                |
| 获奖成果                     | 项目名称、成果登记号、主要完成人、主要完成单位、完成机构所在地、成果水平、奖励名称、获奖年度、奖项等级、奖项、推荐部门、成果简介                    |                |
| 技术需求                     | 技术需求、技术分类、需求期、需求机构、需求地点   |                |
| 农业机构(服务机构、企业)            | 机构名称、机构描述、机构链接、机构类型、机构级别、建立时间   |                |
| 粮食生产数据                   | 作物类型、单位亩产、种植面积、总产量、亩产的增长速度、面积的增长速度、产量的增长速度、面积占粮食比重、产量占粮食比重、亩产、面积占全国平均水平、产量占全国、年份、省份 | 国家农业科学<br>数据中心 |
| 畜禽生产数据<br>(总类、羊、牛、兔、禽、蜂) | 产量、产量占全国比重、年份、地点  |                |

表 2 科技文献元数据结构表

Table 2 Metadata schema of scientific and technological literature

| 资源类型   | 元数据   | 数据来源      |
|--------|---|-----------|
| 专利     | 授权年度、申请号、名称、主分类号、分类号、申请人、发明人、公开日、公开号、专利代理机构、代理人、申请日、地址、摘要、国省代码、专利权人                                 | 国家知识产权局   |
| 中文期刊文摘 | 题名、作者、机构名称、关键词、文摘、出版年、期刊名、卷期号、发表时间、来源、地点  | CNKI      |
| 外文期刊文献 | work_id、题名、其他语种题名、集合题名、机构名称、贡献者机构名称、贡献者名称、贡献者类型、来源名称、出版类型、OA in One 分类号、关键词、替代关键词、语种、文摘卷、期、出版年、出版日期 | OA in One |
| 政策法规   | 题名、机构名称、OA in One 分类号、产业领域、产业链、政策法规类型、发布时间、生效时间等  |           |

中，为了方便管理部门的管理和支持，从管理角度设计了产业链流程。以四川农村科技发展中心为例，设计了适应科技管理的 5 个产业链环节，覆盖了繁育、种养殖、加工、物流及副产物利用（表 3）。该设计与机构管理职能紧密相关，目标指向性明确。但综合考虑到农业科技管理的任务重点及产业分类管理需求，在该产业链划分基础上，本研究对科技管理的产业链环节进行了重构。设计了良种繁育、种养殖、初加工、

精深加工、副产物综合利用和储运物流 6 个环节，将加工环节划分为初加工和精深加工，调整了物流运输的环节排序，并给出定义（表 3、表 4）。

(3) 产业化的分类特征。科技管理数据的另外一个特色是按照产业特色分类，基于产业需求并不要求分类精细化（图 1）。以四川省农村科技发展中心制作的“星火计划项目元数据”为例，在数据的规范化处理上按照四川省的重点发展产业进行划分，这与科技

表 3 农业产业链环节设置一览表

Table 3 The list of agricultural industrial chain

| 时间         | 产业链环节                               |
|------------|-------------------------------------|
| “十二五”“十三五” | 品种繁育——标准化种养殖——加工——物流运输——副产物循环利用     |
| 产业链重构      | 良种繁育——种养殖——初加工——精深加工——副产物综合利用——物流运输 |

表 4 产业链环节级其定义

Table 4 The definition of industrial chain

| 产业环节    | 环节定义   |
|---------|--|
| 良种繁育    | 将通过驯化、驯养或者培育等选育手段得到的优良品种按照一定程序进行扩大繁殖使优良性状得以保存延续并推广于生产的过程                       |
| 种养殖     | 即种植和养殖。种植即农作物的栽培环节, 通过各种种植方式生产农作物的过程。养殖是指培育和繁殖动物农产品的过程, 包括家畜养殖、家禽养殖、水产养殖和特种养殖等 |
| 初加工     | 根据农产品性质和销售运输保管的要求在产地对产品进行的初步加工处理的过程, 起减损、增收、提质、增效作用                            |
| 精深加工    | 经初加工后的农产品进一步精细加工和深度加工, 是促成农产品产生更高附加值的生产活动、工艺或者装备                               |
| 副产物综合利用 | 对农产品在加工过程中产生的副产物(如稻壳、玉米芯、菜叶菜帮、甘蔗渣等)进行再利用的过程, 是提高农产品综合利用的生产活动或相关技术              |
| 储运物流    | 提供农产品仓储、配送、加工、分销、现货交易市场、国际货运代理、进出口贸易、信息等全过程综合物流服                               |

| 项目名称                     | 项目类型   | 产业领域 | 产业链环节       |
|--------------------------|--------|------|-------------|
| 青峪猪繁育及规模化养殖技术推广与示范       | 国家星火计划 | 生猪   | 良种繁育、标准化种养殖 |
| 油用牡丹优良品种推广关键技术集成与示范      | 国家星火计划 | 花卉   | 良种繁育        |
| 秦巴山区冬草莓标准化种植及加工技术集成与示范   | 国家星火计划 | 水果   | 标准化种养殖      |
| 广元市旺苍县红茶无菌化加工技术应用与示范推广项目 | 国家星火计划 | 茶叶   | 精深加工        |
| 块菌保育促繁和高效栽培技术示范          | 国家星火计划 | 食药菌  | 良种繁育、标准化种养殖 |
| 羊肚菌高效生产技术示范与应用           | 国家星火计划 | 食药菌  | 标准化种养殖      |
| 柴胡 GAP 技术集成及其在秦巴山区的示范应用  | 国家星火计划 | 中药材  | 精深加工        |
| 优质蛋鸡智能化健康养殖及综合配套技术应用与示范  | 国家星火计划 | 家禽   | 标准化种养殖      |
| 优质黄肉猕猴桃新品种“金什 1 号”示范与推广  | 国家星火计划 | 水果   | 良种繁育、标准化种养殖 |
| 二次连续式发酵四川泡菜新产品开发与集成示范    | 国家星火计划 | 蔬菜   | 精深加工        |

图 1 星火计划项目元数据截图

Fig.1 The screenshot of spark programs metadata

文献常用《中国图书分类法》《中国科学院图书分类法》的分类方法精细化有明显区别, 其分类级别上类似于农业分类下的二级或三级类目, 较宏观。

(4) 科学数据与科技文献使用特征。管理中常用的数据以项目数据、成果数据、专家人才等数据为主, 以科学观测、实验等为代表的科学数据使用较少, 生产数据有其价值但缺乏规模性使用场景, 期刊论文和专利为代表的科技文献的知识挖掘不多, 政策法规的支撑不足。图书情报学的知识组织和知识挖掘方法在农业科技管理还有进一步应用空间。

### 3.2.2 产业知识服务需求

根据农业科技管理的特征来看, 管理部门的知识服务需求侧重产业宏观和中观管理, 如决策、预测、制定和分析指标、制定政策、目标和战略规划等。和

科学研究的知识服务需求特征不同, 更偏向于宏观、反应整体的数据使用。

本研究的知识服务主体是从事农业科技管理的部门、从事农业产业研究的科研机构 and 大学、企业, 目标用户是科技管理决策人员, 从事产业研究、产业转化的科研人员和企业管理人员。

### 3.3 农业产业知识服务数据生命周期构建

数据生命周期是指科学数据从产生到消亡的全过程。从数据管理角度, 通常将生命周期划分数据采集、数据存储、数据处理、数据传输、数据交换、数据销毁 6 个阶段。在具体的应用环境中, 根据数据使用情况增减环节。如 2018 年国务院办公厅印发的《科学数据管理办法》中涉及的科学数据管理生命周期是数据

采集和交汇、数据保存、数据共享利用、数据保密安全 4 个方面。本研究按照科学数据管理中参与产业知识服务的数据，按照信息管理生命周期流程和农业产业管理特征，初步构建农业产业管理科学数据生命周期（图 2）。

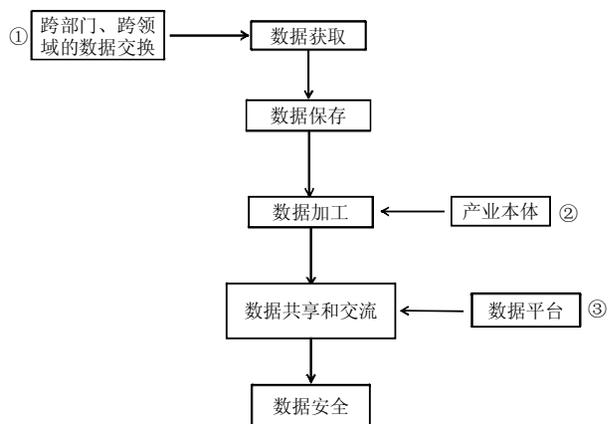


图 2 农业产业管理科学数据生命周期模型

Fig.2 Scientific data life cycle model of agricultural industry management

该生命周期有 3 个关键环节，一是跨部门的数据交换，通常有部门之间的协商交换和开放数据获取两种方式；二是根据知识服务和数据的特征构建产业特征的 ontology，实现非结构化数据的语义关联；三是基于数据平台的共享交流，和第一个环节相比，有知识产权的制约，此外数据完整性也有一定损失或影响。综上，研究将依据农业科技管理特征和需求，构建产业 ontology，进而实现数据之间的关联。

### 3.4 农业产业 ontology 模型构建

#### 3.4.1 本体构建原则

本体模型是产业知识服务模型中的关键环节之一。作为产业知识服务的技术支持，本体可以帮助建立科学数据之间、科学数据与科技文献数据之间的关联。本研究中的农业产业 ontology（以下简称 ontology），是介于领域 ontology 和科学 ontology 之间的简单应用 ontology。按照 MC Guinness 对简单 ontologies 的定义，将包括 3 个内容：一是建立有限词汇表；二是概念和术语之间有关系解释；三是类之间有严格分层的子类关系。

#### 3.4.2 本体构建流程

本研究采用七步法构建 ontology。在确定专业领域和数据类型后，考查复用 ontology，构建了概念、属性、关系和实例。由于是基于领域的 ontology 研究，研究中邀请了专家介入，辅助控制质量。构建流程参见图 3。

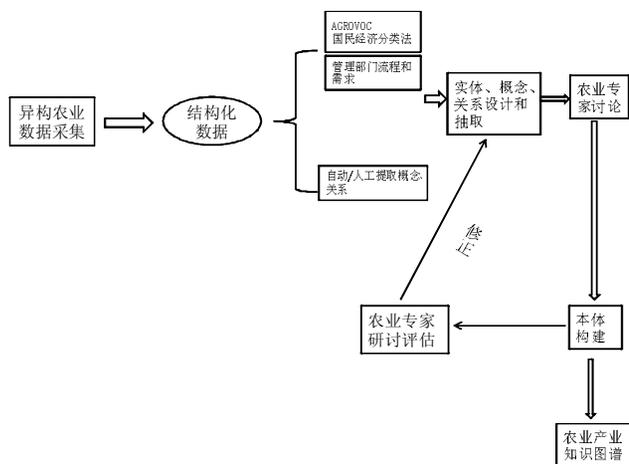


图 3 农业产业 ontology 构建流程图

Fig.3 Flow chart of agricultural industry ontology construction

#### 3.4.3 本体模型

(1) 本体语言及结构。农业产业 ontology 包括概念、属性、和实例。本体主要分为对概念层与实例层的构建。概念层构建包含对类、类的层级结构、类间关系与属性的定义，形成概念级别的语义网络。实例层主要是在概念层的基础上对具体实例进行描述，将实例填充到概念语义网络中，实现具体知识的组织与应用。

(2) 本体概念构建。首先分析了科技产业管理要素和 R&D 活动评价指标，创建概念结构；其次调研了领域相关的 ontology、叙词表和分类法，吸收其中的产业特性；然后根据农业核心数据分析概念，搭建类目层级结构；最后在专家建议和数据实测基础上完善概念。研究参考了农业多语种叙词表（AGROVOC）、科技知识组织体系（STKOS）、《中华人民共和国农业技术推广法》、国民经济分类法、中国图书分类法、中国科学院图书分类法，最后形成 13 个一级类目、65 个二级子类、若干三级子类的农业产业 ontology 概念层级框架，层级深度最深为三级，概念内容覆盖了产业环节分类、农业领域分类、产业投入和产出等领域（图 4）。

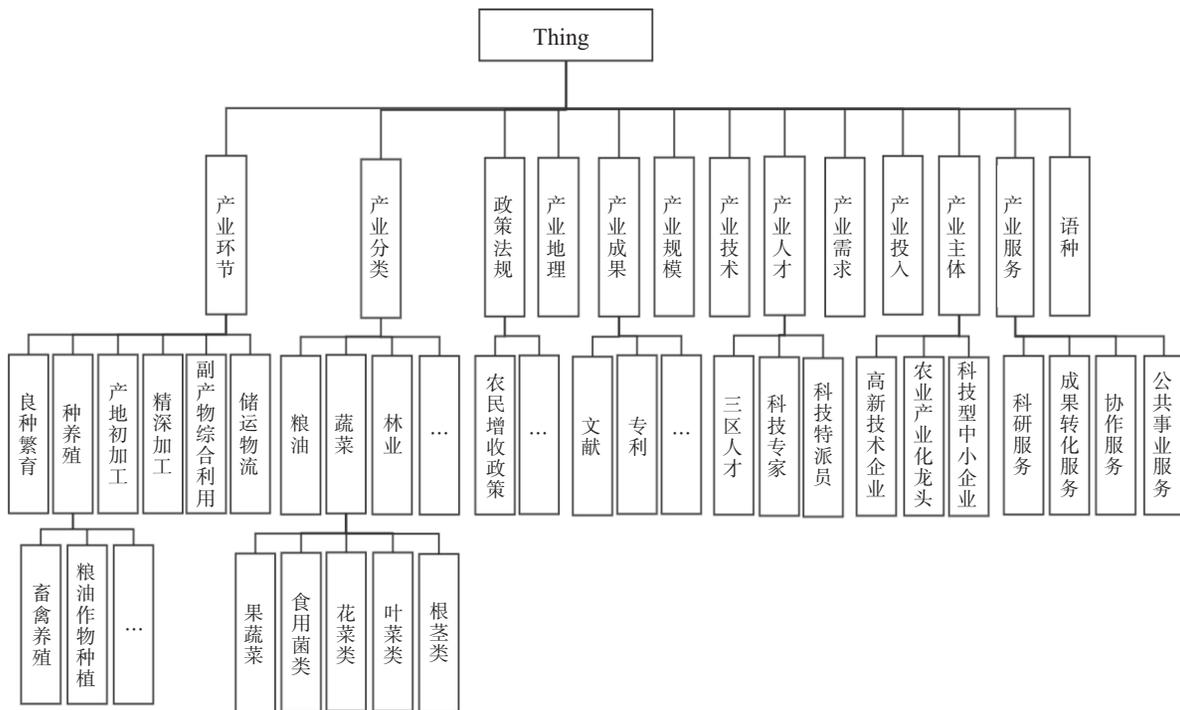


图4 概念层级效果示意图

Fig.4 Concept hierarchy of the ontology

(3) 关系设计。关系设计部分主要厘清并构建了概念之间的关系，包括基本关系和自定义关系。基本关系复用了RDFS、DC等描述框架的一些常用关系，对于无法被覆盖的其他特殊关系，采用自定义关系进行描述。

(4) 本体属性。属性分为对象属性和数据属性。对象属性主要用于描述两个概念之间的关系。在完成关系设计并交予领域专家审核后，这些关系在本体构建中，以对象属性的形式被构建。比如“产业成果”-“有关于”->“产业环节”。数据属性描述了该概念的非关联性属性，以属性值的方式表达，不与其他概念产生联系。比如产业主体的机构名称、文献的标题。

#### 3.4.4 数据测试与修正

本体设计的核心工作之一是根据科学数据和科学文献的外部性和内部共性特征建立5个概念：产业环节、产业分类、产业技术、产业地理和政策法规。这5个概念能否带入科学数据和科技文献将影响数据关联准确性。

测试部分以专家讨论和数据实践为主，分4步完成。首先，在初稿基础上邀请农业科学、农业管理和

图书情报领域专家进行本体讨论，并根据专家意见修改完善；其次邀请农业科学和图书情报领域的研究人员进行数据验证；第三在平台中进行数据关联验证；最后修正本体。测试人员从描述准确性、区域特征和可操作性对26190条数据进行了语义提取、分析和修正，获得有效数据23100条，语义提取率达到88.20%。测试认为本体实测效果较好，总体可以满足科学数据和科技文献数据的关联需求，兼顾了科学数据的宏观和微观描述，符合四川省农业产业特色，分类操作过程简单、易用、易标引，概念准确、无歧义。

测试发现主要有3个问题：一是语义粒度粗细差异。以产业环节为例，农业政策法规、农业项目、农业机构数据的分类较为宏观，一条数据可能覆盖多个三级类目，或者二级类目。修正后，将宏观和中观相关数据进行二次测试，分类以一级或二级类为主，能够实现粗的语义对齐。二是概念覆盖不全。通过专家研讨和测试两个环节补充了内容，如经专家建议补充了四川省农产品、林木、药材等特色产业分类。三是分类视角问题。测试人员以农业经济管理专业的硕、

博士研究生、图书馆学硕士研究生为主,优点是测试流程规范、语义识别专业度较高,缺点是跨领域工作增加了语义提取难度,后期通过第三方审核和交叉审核,能够及时发现和矫正。

## 4 研究实践

### 4.1 Vocbech3 本体构建实践

本部分主要是使用四川农业产业管理相关的科学数据和科技文献,在该软件中进行实践,检验本体结构的合理性和可行性。构建操作分为3步。

(1) 搭建类目层级。首先将概念进行分类与组织,明晰概念及基础关系,归纳形成明确的类目,搭建类目层级结构。类目层级主要通过 `rdfs:subClassOf` 属性来实现,若类 A 与类 B 有“A-subClassOf->B”的类间关系,则说明 A 类与 B 类有层级关系,且 A 是 B 的子类。类目创建界面如图 5 所示。

(2) 添加属性。根据本体设计部分的关系设计,

结合实例数据的元数据描述,通过归纳、转化、整合复用得出本体所需的属性。之后,划分对象属性与数据属性,对属性的定义域与值域进行进一步明晰,得到清晰完整的属性清单,并在 Vocbench3 中进行构建。根据整理好的属性清单,创建对象属性及数据属性,并在属性设置窗口中进一步定义属性的值域与定义域。属性创建界面如图 6 示。

(3) 实例层构建。Vocbench3 中创建实例需要先在 Class 视窗中选中需要创建实例的类目,再在 Instance 视窗中,点击创建按钮构建实例,在 instance 编辑窗口中添加实例相应的属性,完成实例构建。实例创建界面如图 7 所示。创建完成的实例可以通过可视化方式展示,如图 8 所示。

### 4.2 基于系统的数据关联实践

#### 4.2.1 农业产业知识服务平台

知识服务平台将提供多类型科学数据检索服务,检索结果可以按照产业要素关联和聚类,进而提高用户获取隐性知识的速率,扩展产业知识,启发新需求。

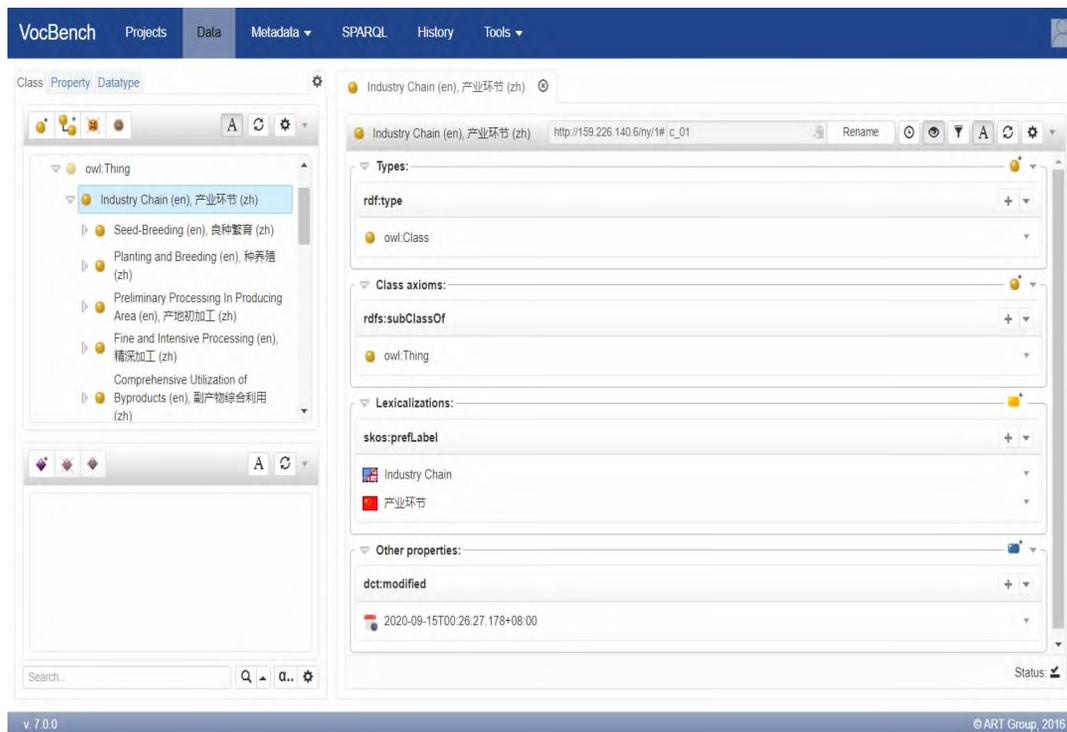


图 5 Vocbench3 类目创建示意图

Fig.5 Class construction in Vocbench3

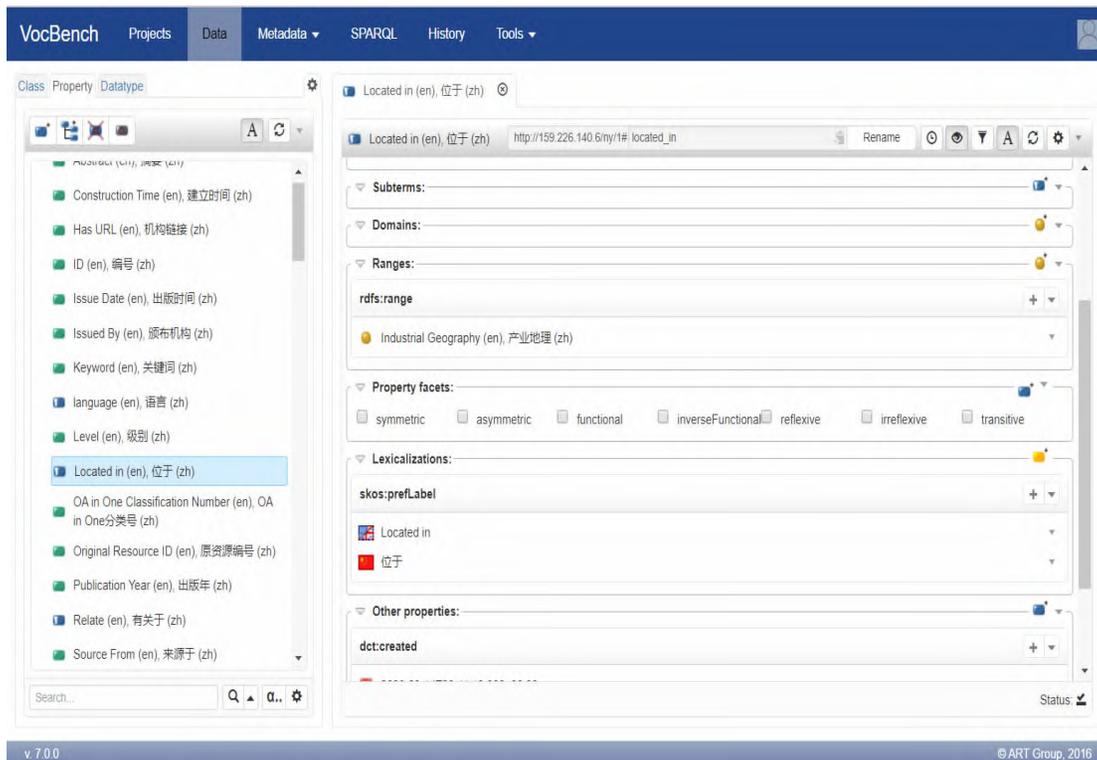


图 6 Vocbench3 属性创建示意图

Fig.6 Property construction in Vocbench3

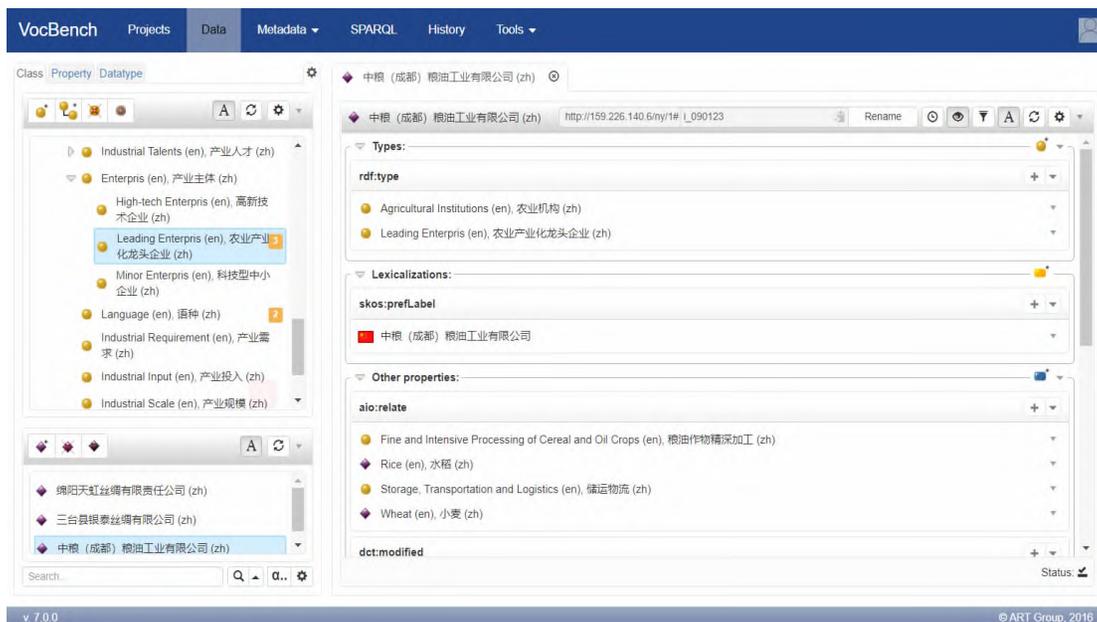


图 7 Vocbench3 实例创建示意图

Fig.7 Instance construction in Vocbench3

平台基于 Java 语言开发, 数据库采用 Mysql5.5, 服务器 Tomcat8.5。整个系统采用前后端分离开发, 系统部署于云环境, 采用 Zstack Iaas 平台创建和管理系统业

务云主机采用扁平网络搭建安全的网络结构。从数据关联情况来看, 可以通过产业领域、产业环节、资源类型、语种和地点进行分面聚类。平台选取产业要素



等作为建设目标的平台不多, 服务力度不够。因此, 本研究结合四川省农业发展特点以及发展方向, 在主平台下建立四川省农业领域专家信息与服务平台。目的是搜集和展示相关领域专家信息, 并采集农业专家公开发表的论文信息、农业领域政府机构和专业研究机构的新闻信息。目前平台内国内专家的数量为 859 条, 国外专家的数量为 325 条。主平台和子平台数据通过产业本体中的“产业人才”进行关联。

## 5 研究成果

研究根据科学数据特征和知识服务需求, 将本体概念和产业要素结合起来, 提出农业产业知识服务模式。并就产业知识服务模式中的两个关键要素: 产业本体和数据平台建设进行研究。研究建立了科学数据和科技文献关联的本体模型并在 Vocbench3 中实现, 构建了农业知识服务平台和农业专家子平台, 在系统中实现了 24 284 条数据的产业知识共现, 参与数据包括 16 类农业科学数据和 4 类科技文献。

研究认为来自农业管理机构的数据和来自情报机构的科技文献, 在信息描述粒度和层次上有所不同, 呈碎片化分布在产业链各环节。通过构建农业产业管理本体, 能把多源异构数据关联融合于产业管理逻辑下, 帮助用户知识发现。但是融合后数据价值还有待挖掘。

## 6 研究局限与展望

研究基本实现农业科技知识发现的服务目标, 探索了农业科学数据在科技管理的应用, 提出了数据关联方法, 然而由于数据总量较少, 还需要进一步充实和展示关联, 实现实例数据的批量转化, 以及后端的知识图谱开发与应用。本体模型覆盖了 13 个概念, 概念相对庞杂, 将进一步修正, 如把产业地理调整为属性值。此外, 将在现有基础上挖掘数据链的价值, 开展全产业链视角下的技术预见理论和方法研究。

### 参考文献:

- [1] 任俊为. 知识经济与图书馆的知识服务[J]. 图书情报知识, 1999(1): 3-5.  
REN J W. Knowledge economy and library knowledge service[J]. Documentation, information & knowledge, 1999(1): 3-5.
- [2] 张晓林. 走向知识服务: 寻找新世纪图书情报工作的生长点[J]. 中国图书馆学报, 2000(5): 30-35.  
ZHANG X L. Towards knowledge service: Seeking development opportunities for library and information services in the 21st century[J]. Journal of library science in China, 2000(5): 30-35.
- [3] 李霞, 樊治平, 冯博. 知识服务的概念、特征与模式[J]. 情报科学, 2007(10): 1584-1587.  
LI X, FAN Z P, FENG B. The concept, characteristics and model of knowledge-intensive services[J]. Information science, 2007(10): 1584-1587.
- [4] 柯平, 邹金汇. 后知识服务时代的图书馆转型[J]. 中国图书馆学报, 2019, 45(1): 4-17.  
KE P, ZOU J H. Library transformation in the post-knowledge service era[J]. Journal of library science in China, 2019, 45(1): 4-17.
- [5] 刘健宏, 池敏青, 吴越. 大数据背景下农业分学科知识服务模式研究[J]. 台湾农业探索, 2018(3): 77-81.  
LIU J H, CHI M Q, WU Y. Knowledge service model of agricultural discipline under large data background[J]. Taiwan agricultural research, 2018(3): 77-81.
- [6] 王丹阳. 农业知识服务模式探究[D]. 北京: 北京印刷学院, 2019.  
WANG D Y. Research on knowledge service model of agriculture[D]. Beijing: Beijing institute of graphic communication, 2019.
- [7] 柏永青, 杨雅萍, 孙九林. 国内外科学数据管理办法研究进展[J]. 农业大数据学报, 2019, 1(3): 5-20, 4.  
BO Y Q, YANG Y P, SUN J L. Advances in the study of domestic and foreign scientific data management methods[J]. Journal of agricultural big data, 2019, 1(3): 5-20, 4.
- [8] 陈娉婷, 官波, 沈祥成, 等. 大数据时代开放式农业信息知识库构建研究[J]. 东北农业科学, 2018, 43(5): 60-64.  
CHEN P T, GUAN B, SHEN X C, et al. Studies on the construction of open repository of agricultural information in big data era[J]. Journal of northeast agricultural sciences, 2018, 43(5): 60-64.

- [9] SGOUROPOULOU C, KOUTOUMANOS A. Building metadata architectures: A case for e-research[J]. International journal of metadata semantics & ontologies, 2014, 9(4): 275-288.
- [10] AYDIN S, AYDIN M N. Semantic and syntactic interoperability for agricultural open-data platforms in the context of IoT using crop-specific trait ontologies[J]. Applied ENCES, 2020, 10(13): 1-27.
- [11] WISNUBHADRA I, ADITHAMA S P, BAHARIN S S K, et al. Agriculture spatiotemporal business intelligence using open data integration [C]//2019 international seminar on research of information technology and intelligent systems(ISRITI), 2019.
- [12] ZHANG M, LI Z, LI F. Discussion on key technologies in forestry fundamental scientific information cloud service platform[C]//Green computing & communications, IEEE, 2013.
- [13] MARTIN C, CADIOU C, EMMANUELLE JANNES-OBER. Data management: New tools, new organization, and new skills in a French research institute[J]. LIBER quarterly, 2017, 27(1): 73-88.
- [14] ENAYAT, RAJABI, SALVADOR, et al. A linked and open dataset from a network of learning repositories on organic agriculture[J]. British journal of educational technology, 2015.
- [15] FERRAG M A, SHU L, YANG X, et al. Security and privacy for green IoT-Based agriculture: Review, blockchain solutions, and challenges[J]. IEEE access, 2020, 99.
- [16] 科学数据共享工程[EB/OL]. [2021-10-08]. [http://www.most.gov.cn/ztzl/kjzg60/kjzg60hhcj/kjzg60jcyj/200909/t20090911\\_72832.html](http://www.most.gov.cn/ztzl/kjzg60/kjzg60hhcj/kjzg60jcyj/200909/t20090911_72832.html).  
The construction strategy of national scientific data sharing program[EB/OL]. [2021-10-08].[http://www.most.gov.cn/ztzl/kjzg60/kjzg60hhcj/kjzg60jcyj/200909/t20090911\\_72832.html](http://www.most.gov.cn/ztzl/kjzg60/kjzg60hhcj/kjzg60jcyj/200909/t20090911_72832.html).
- [17] 贺玲玉. 基于关联数据的农业信息资源整合研究[D]. 武汉: 华中师范大学, 2014.  
HE L Y. Research on the integration of agricultural information resources based on the linked data[D]. Wuhan: Central China normal university, 2014.
- [18] 王剑, 黄朝光. 海量农业科学数据存储体系架构与方法研究[J]. 广东农业科学, 2015, 42(2): 152-156.  
WANG J, HUANG C G. Research on storage architecture and method for big agricultural scientific data[J]. Guangdong agricultural sciences, 2015, 42(2): 152-156.
- [19] 陆丽娜, 王萍, 于啸. 农业科学数据监管平台构建研究[J]. 图书情报工作, 2017, 61(10): 68-73.  
LU L N, WANG P, YU X. Research on the construction of agriculture data curation platform [J]. Library and information service, 2017, 61(10): 68-73.
- [20] 赵瑞雪, 李娇, 张洁, 等. 多场景农业专业知识服务系统构建研究[J]. 农业图书情报学报, 2020, 32(1): 4-11.  
ZHAO R X, LI J, ZHANG J, et al. Construction of multi-scenario agricultural knowledge service system[J]. Journal of library and information science in agriculture, 2020, 32(1): 4-11.
- [21] 赵华, 王健. 科学数据出版现状及对中国农业科学数据出版的启示[J]. 农业展望, 2016, 12(8): 53-57.  
ZHAO H, WANG J. Circumstance on scientific data publication and its inspiration to China's agricultural scientific data publication[J]. Agricultural outlook, 2016, 12(8): 53-57.
- [22] 彭秀媛, 王枫, 周国民. 面向重用的农业科学数据共享模式研究[J]. 农业经济, 2019(1): 87-89.  
PENG X Y, WANG F, ZHOU G M. Research on reuse-oriented data sharing model for agricultural science[J]. Agricultural economy, 2019(1): 87-89.
- [23] 科学技术[EB/OL]. [2020-07-29] [http://www.stats.gov.cn/tjsj/zbjzs/201912/t20191202\\_1713041.html](http://www.stats.gov.cn/tjsj/zbjzs/201912/t20191202_1713041.html).  
Science and technology[EB/OL]. [2020-07-29] [http://www.stats.gov.cn/tjsj/zbjzs/201912/t20191202\\_1713041.html](http://www.stats.gov.cn/tjsj/zbjzs/201912/t20191202_1713041.html).
- [24] TONY H, STEWART T, KRISTIN T.《第四范式: 数据密集型科学发现》[M]. 北京: 科学出版社, 2012  
TONY H, STEWART T, KRISTIN T. The fourth paradigm: Data-intensive scientific discovery[M]. Beijing: Science press, 2012
- [25] 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  
HUANG X J. Link study of scientific data and scientific literature based on content features[J]. Journal of modern information, 2018, 38(1): 56-59
- [26] 鲜国建. 农业科技多维语义关联数据构建研究[D]. 北京: 中国农业科学院, 2013.  
XIAN G J. Research on construction of multidimensional semantic linked data of agricultural science and technology[D]. Beijing: China Academy of Agricultural Sciences, 2013.

nese academy of agricultural sciences, 2013.

- [27] 李帆. 面向高铁的数据压缩及趋势关联分析算法研究[D]. 西安: 西安理工大学, 2016.

LI F. Research on data compression and trend association analysis algorithm for high-speed train[D]. Xi'an: Xi'an university of technology, 2016.

- [28] 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.

ZHANG X, WEN Y, YANG N, et al. A citation-based association algorithm between scientific literature and data: A case study of high energy physics[J]. Information studies: Theory & application, 2019, 42(10): 151-156.

- [29] 丁文姚, 李健, 韩毅. 我国图书情报领域期刊论文的科学数据引用特征研究[J]. 图书情报工作, 2019, 63(22): 118-128.

DING W Y, LI J, HAN Y. Research on the characteristics of scientific data citation in journal articles in library and information science in China[J]. Library and information service, 2019, 63(22): 118-128.

- [30] 柴苗岭, 黄琳, 任运月. 重要开放农业科学数据资源建设现状综

述[J]. 农业图书情报学报, 2020, 32(10): 25-34.

CHAI M L, HUANG L, REN Y Y. A review of construction of major agricultural open scientific data resources[J]. Journal of library and information science in agriculture, 2020, 32(10): 25-34.

致谢:

研究在本文框架构建过程中得到中国农业科学院农业信息研究所张学福和潘淑春研究员的指导和支  
持。本文设计得到四川农业大学农业经济学博士生导师王芳教授的支持, 在 Vocbench3 概念、属性、关系和实例构建中得到 Armando Stellato 教授、姜恩波研究员的支持, 数据测试部分得到史俊晖博士的支持。在此表示衷心感谢!

感谢国家科技基础条件平台——国家农业科学数据中心 (<http://www.agridata.cn>) 为本研究提供科学数据支撑。感谢中国科学院文献情报新增能力建设项  
目——“开放知识资源体系建设” (<http://oa.las.ac.cn>) 为本研究提供外文论文和政策法规数据支撑。

## Research and Practice on Association of Scientific Data and Scientific Literature Oriented to Knowledge Service of Agricultural Industry

CHAI Miaoling<sup>1,2,3</sup>, ZOU Yixing<sup>4</sup>, TAN Rongzhi<sup>3\*</sup>, ZENG Yi<sup>5</sup>, REN Yunyue<sup>6</sup>

(1. Chengdu Documentation and Information Center, Chinese Academy of Sciences, Chengdu 610041; 2. School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190; 3. TAN Rongzhi, Key Laboratory of Mountain Hazards and Earth Surface Process, Chinese Academy of Sciences, Chengdu 610041; 4. Sichuan Rural Science and Technology Development Center, Chengdu 610042; 5. Business school, City University of Hong Kong, Hong Kong 999077; 6. School of Public Administration, Sichuan University, Chengdu 610065)

**Abstract:** [Purpose/Significance] In order to solve the problems of incomplete data chain and insufficient data support decision-making ability in agricultural science and technology management, this paper put forward a set of industrial knowledge service model based on the needs of scientific management in agriculture, and completed the key part of the model, the research and practice on association of

scientific data and scientific literature. This study will provide theoretical and practical cases for library and information science (LIS) research institutions to serve agricultural science and technology management and industrial development. [Method/Process] Firstly, this paper reviews the current situation of research into knowledge service and data association. Secondly, based on the analysis of characteristics of industrial requirements and data management life cycle in agricultural science and technology management, the paper constructed a knowledge service model with industrial characteristics. Then, based on the agricultural science data and scientific literature of Sichuan Province, the study designed and constructed an ontology model and system based on industrial elements. [Results/Conclusions] 20 data types, 24 284 scientific data entries and related scientific literature were integrated. And we established multi-source heterogeneous data association in 1 system platform and 1 sub-platform with Chinese and English data. The research results will provide reference for cross-field cooperation between agricultural science and technology management and LIS research, and cross-departmental sharing of scientific data and scientific literature.

**Keywords:** agricultural industry; scientific data; scientific literature; data association; ontology; industrial chain; big data