

# 学者身份识别的机制及关键技术研究\*

常 娥

**摘 要** 文章从探讨学者身份识别的概念、功能和内涵出发,深入剖析人名规范数据库和学者身份唯一标识符系统这两种典型解决方案,揭示了学者身份识别的内在机制和关键技术,以期为后续学者身份识别研究提供参考和借鉴。

**关键词** 学者身份识别 规范控制 名称歧义 唯一标识符

**引用本文格式** 常娥. 学者身份识别的机制及关键技术研究[J]. 图书馆论坛, 2015 ( 10 ): 88- 95.

## The Mechanism and Key Technology of Scholar Identification

Chang E

**Abstract** This paper discusses the concept, function and meaning of scholar identification, then deeply analyzes the authority control database of author name and the scholar unique ID system, which are two kinds of the classic solutions to scholar identification. Finally, it brings insight into the mechanism of scholar identification and research on the key technology to identify scholars.

**Keywords** scholar identification; authority control; name ambiguity; unique ID

学者身份识别是信息组织与检索、文献计量、科研评价等领域长期关注的热点问题。如今学者身份识别问题正突破传统图书情报领域的研究视角,引起出版商、基金组织、学术社团等机构的极大关注。近年国外大型图书出版集团组织推出可嵌入科学研究生命周期的学者唯一身份识别符,再次引发学界关于学者身份识别研究的热潮。此外,随着在线学术交流模式的形成和替代计量学的发展,学者身份识别问题变得愈发重要。然而国内文章大多是关于 ORCID、ResearcherID 等科研人员唯一标识符的引荐,对于学者身份识别的内在机制和关键技术缺乏深入思

考。本文从探讨学者身份识别的概念、功能和内涵出发,深入剖析人名规范数据库和学者唯一身份识别符系统这两种典型解决方案,揭示学者身份识别的内在机制和关键技术,以期为后续学者身份识别研究提供参考和借鉴。

### 1 学者身份识别的概念

学者指具有一定专业技能、知识水平、创造能力,能在相关领域表达思想、提出见解、引领社会文化潮流的人<sup>[1]</sup>,有时被称为专家、科研工作者、研究人员等。从身份定位角度来说,学者身份是指学者不同于他人的特殊个性,其身份由

\* 本文系国家社会科学基金项目“图书馆资源组织中的数据关联机制研究”(项目编号:14CTQ005)和中央高校基本科研业务费项目“东南大学科研人员标识(SEU-AUTHOR)系统研究”(项目编号:2242015S2007)研究成果之一;本文得到国家留学基金资助

学者姓名、学科领域和专业级别等信息加以确定。其中姓名是确认学者身份的主要标识,学科领域和专业级别等是附加标识。在学术界,当我们提及某一专家时,首先需确认的是该专家的姓名,然后才是他的学科领域、专业级别、工作单位,以及年龄、性别、荣誉称号等信息。

学者身份识别是指通过一定的方法和技术来识别学者身份,达到准确区分每一位学者的目的。就识别内容而言,学者身份识别有广义和狭义之分。狭义的学者身份识别是指将学者姓名与其名称所有版本统一起来,可简称学者姓名识别;广义的学者身份识别不仅需要将学者姓名进行规范化处理,而且需附加学者性别、出生年月、工作单位、学科领域等相关信息,从而获得某一学者完整的身份信息。一般将相对精确、固定的附加信息称为自然属性,包括学者性别、出生日期、死亡日期、出生地等。将含混、易变的附加信息称为社会属性,包括学者工作单位、学科领域、专业级别、所获荣誉等。

日常工作、生活和学习交流环境中,学者姓名是学者身份信息的统一指代,这一点毋庸置疑。由于学者个人名称复杂,同名和异名等情况屡见不鲜,增加了学者身份辨别的难度。学者身份识别的关键是研究如何将学者姓名及其名称的所有版本统一起来,建立唯一标识符。下文将分析学者异名和同名的发生机理以透视学者名称的复杂性,并剖析学者身份识别的内涵和功能,为揭示学者统一身份标识符的内在机制奠定基础。

## 2 学者个人名称的复杂性

### 2.1 学者异名

学者异名即同一学者拥有各种不同的名称。为便于论述,本文将学者名称区分为本名和异名两种基本形式。本名指学者个人社会性身份证件上的真实姓名,除此之外的其他一切为人所知的称呼视作异名,包括全称、简称、昵称、网名、笔名、缩写名、中文汉语拼音等<sup>[2]</sup>。为什么会出现学者异名?异名是学者主动添加或更改而获得

的,主要原因有三点:(1)学者在职业发展生涯中正式改换姓名,使得科研成果在改名前后有了不同署名;(2)学者在不同语种的期刊中发表研究成果,采用本名的音译、缩写或汉语拼音等方式进行署名,因而出现了异名;(3)学者交流的环境和偏好发生变化,在线科研交流环境中,学者申请使用各种网名、昵称和笔名等。

不同情形产生的学者异名,在字面上与本名关联的程度不同,计算机识别的难度也不尽相同。第一种,学者改名而产生的异名,一般会延续原姓氏,与本名间存在一定字面联系,但难以自动识别;第二种异名比较普遍,与本名间不存在字面关联,但属于不同语种间的翻译和转换,借助计算机技术可自动识别;第三种异名情况较复杂,与本名间即不存在字面关联,又不属于不同语种间的翻译和转换,很难利用计算机技术进行自动识别。因此,笔名、网名、昵称等增加了学者异名现象的复杂性和处理难度。

### 2.2 学者同名

学者同名即不同学者拥有相同的名称。当把不同学者的科研成果集中起来形成数据库时出现同名异人现象,难以确认科研成果的精确归属。并且随着数据库规模的扩大,同名出现频率随之增高,即同一姓名关联到的不同学者越多,这是学者同名现象出现的根本原因。据此推理,如果将学者不同语种的出版物,以及学者的学术博客、微信、专业论坛帖等成果记录一并考虑,纳入更加庞大的学者研究成果库,由于简称、缩写名、译名、网名等各种异名的存在,将极大增加学者异名出现同名的概率,从而增加学者身份辨别的难度。

过去图书馆、数据库商等机构在组织信息资源的过程中,一般是在各自数据范围内针对学者本名出现的同名现象进行区分和控制,较少考虑学者异名出现的同名现象。由于同名是学者被动接受的,且随着数据源的变化而变化,因此综合不同数据源,集中规范控制学者本名和异名出现的同名现象是未来信息组织领域非常值得研究的方向。

### 3 学者身份识别的功能和内涵

如今学术交流圈已由图书、期刊等传统出版领域,扩散到微博、博客、学术论坛等网络数字环境中。学术交流网络化是趋势,在线学术交流这一全新学术交流模式正在生成。在线学术交流不仅反映了网络时代科研工作者交流偏好的变化,更重要的是提高了科研交流的效率和受众面。学术交流模式的改变使得科研成果不再局限于完整独立的学术论文或著作,学者们不断将学术成果中的部分观点、结论、图表或实验记录与数据等,以博文、微博等方式在网络中进行发布和交流。因此,本文认为学术记录一词更能全面概括各种碎片化、网络化的科研成果。学术记录除包含论文、专著、专利等传统研究成果形式外,还包括网络上发布的各种学术成果,如软件、程序代码、实验数据、研讨视频、会议ppt、学术博客、学术意见、公众科普等。碎片化学术记录虽然传播共享速度快,受众面更广,但由于格式不规范、内容不够完整,导致其可辨识度、可参考和引用性都很差。要将这些记录都关联到同一学者身上几乎不可能,因而也无法引用,更谈不上分享<sup>[3]</sup>。

学者身份识别的目标是解决学术文章中的作者姓名歧义,以准确区分每一位学者。汇集、区分和引导是学者身份识别的三大基本功能<sup>[4]</sup>:(1)汇集同一学者的不同署名的所有学术记录;(2)区分同名学者的学术记录,辨别学术记录的精确归属;(3)在学者名称间建立参照关系以帮助用户准确获得所需信息。然而将学者姓名与其名称所有版本统一起来,形成统一身份识别符,其内在价值远高于此。试想,如果将学者统一身份识别符嵌入学者学术研究和科研活动生命周期中,在学者和其各种学术记录之间创建关联,那么这种将学者和其各种学术成果关联、引用的能力,可帮助我们更好地了解一个研究人员的学术贡献,改变我们对知识生产的观点,并有可能改变研究和学术活动的内容和开展方式,以及认可和奖励的机制<sup>[5]</sup>。学者们正在探索以文献为基础,以人物为索引的新型知识组织模式,并认为建立

以人为核心的知识关联网络是未来文献支持科研服务及学术生产力分析的重要方向<sup>[6]</sup>。

此外,学者身份识别研究还将极大促进替代计量学(Altmetrics)的发展。替代计量学正在改变以传统文献计量学为核心的科研评价体系,计量数据源也由传统的期刊论文数据库扩展到了学术博客、微博、学术分享和评价、专业问答、开放存取、预印本系统等诸多在线交流平台<sup>[7]</sup>。因此将学者的各种网络名称,例如微信名、博客名、社区昵称、QQ昵称等统一起来,从而将形式各异的学术记录汇集到某个学者名下,是替代计量学发展过程中迫切需要解决的问题。

### 4 学者身份识别的典型解决方案

学者统一身份标识,又可称为学者统一身份标目,或学者身份唯一标识符,可以是一串数字、可以是数字与字母的组合,也可以由学者姓名附加其他信息构成等。不同的机构构建了不同的统一标识符系统,目前以图书情报机构构建的人名规范数据库和国际知名出版集团推出的学者身份唯一标识符这两种解决方案最具代表性。

#### 4.1 人名规范数据库

学者身份识别是一个长期困扰图书馆等信息机构的难题。图书馆界一直重视作者个人名称消歧工作,以个人名称统一标目的规范控制方式,构建了规模庞大的个人名称规范数据库。美国国会图书馆早在20世纪70年代就建立了个人名称规范数据库。2003年由OCLC牵头建立的虚拟国际规范档(VIAF)是国际上最具影响力的名称规范数据库,截至2013年,该库收录个人名称规范数据1800多万条<sup>[8]</sup>。VIAF已开放为关联数据,提供免费下载服务,并与维基百科、国际标准标识符(ISNI)、社会网络与档案文本(SNAC)等数据集深度关联和融合,被认为是万维网利用图书馆规范数据的门槛,在国际数据交换生态系统中扮演重要作用。1995年中国国家图书馆开始个人名称规范工作,从《中国大百科全书》《中国专家大辞典》等各种权威工具书中采集数据,后来又以国图中文普通图书书目数据314字段的责任者小传中的信息作为补充,建成中文名称规范数据

库<sup>[9]</sup>。截至 2009 年底,该库数据共计 83 万多条,其中个人名称规范数据达 76 万多条<sup>[10]</sup>。2003 年由国家图书馆(NLC)、中国高等教育文献保障系统(CALIS)、香港大学图书馆联席会(HKCAN)和台湾汉学研究中心(CCS)联合共建“中文名称规范联合数据库”,收录规范数据记录 88 万多条。

简单地说,图书馆界实施的名称规范控制是使同一实体使用相同、唯一的标题,由这些唯一的标题贯穿整个书目数据库<sup>[11]</sup>,同一实体主要包括作者、书籍、系列图书或合著图书等。图书馆建立个人名称规范数据库的原则是,只要有其作品就为其作名称规范<sup>[12]</sup>,所以个人名称规范数据庞大,个人名称统一标目需要有附加信息进行限定,才具有区分能力。然而选择什么内容对个人名称进行限定目前没有统一标准<sup>[13]</sup>,依然仁者见仁,各有优劣,但生卒年已成为国际通用限定属性。如果个人名称统一标目在有生卒年且重复的情况下,则需使用学科、职称等附加成分进行著录。中国国家图书馆规定附加成分的内容包括:生卒年、职业、学科、职称、国别、外文名、性别、朝代、民族等。在最新的《资源描述与检索》(RDA)中则明确规定了识别作者个人信息的 17 种属性信息,包括个人名称(首先名称、变异名称)、与个人相关的日期(出生日期、死亡日期、个人活跃期)、个人头衔、名称的更完整形式、与个人相关的其他标识、性别、出生地、死亡地、与个人相关的国家、居住地、个人地址、隶属机构、个人语言、个人活动领域、职业或工作、传记信息、个人标识符<sup>[14]</sup>。由此可见,为了使个人名称统一标目具有区分能力,可附加的信息非常复杂,易造成不规范问题。

#### 4.2 学者身份唯一标识符

近年来,学者身份识别问题突破了图书情报机构的研究视角,引起了全球范围内的资源出版商、平台服务商、公益组织及联盟机构的关注,他们推出了与图书馆界完全不同的解决方案,例如 Elsevier 的 Scopus ID,ACM 的 author ID,汤森路透的 Researcher ID、国际标准组织的 ISPI、JISC 的 Name Project、以及 OpenID

等<sup>[15]</sup>。这些解决方案的共同之处在于,它们致力于借助编号系统为学者建立全球性唯一标识符,并试图从学术出版的源头控制学者名称歧义问题。虽然这些解决方案大多具有全球性,但有些并未在学术界极其相关领域得到广泛支持,因为它们或被单一的组织机构所拥有,或因资金缺乏未得到长期维护。

目前开放研究者与贡献者身份(Open Researcher and Contributor ID, ORCID)是解决方案中最具影响的一个。与其他学者唯一标识解决方案相比,ORCID 的优势主要体现在四方面:(1)ORCID 建立学者身份唯一标识符的开放注册系统,不限制学科、机构和地理范围。因此,它可以在全球范围内准确识别每一位学者;(2)ORCID 拥有广泛联盟基础,目前联合出版商、平台商、大学图书馆、学术团体、基金组织、以及国际知名期刊等 250 多个机构,其中 Elsevier、ACM、汤森路透、PubMed、Nature、康奈尔大学图书馆、麻省理工大学图书馆等均是其会员单位;(3)ORCID 不以盈利为目的,对于存储数据和个人使用永久免费,但对会员机构会收取一定的会员费,以维持项目的正常管理和运行<sup>[16]</sup>;(4)ORCID 是一个开放系统,可与其他学者唯一标识符系统进行整合。2012 年 11 月完成与 ResearcherID 的深度整合,建立两个系统间学者身份唯一标识及科研数据的实质性关联,接下来将进行与 Scopus ID 的对接和整合,以实现知识的增值与发现服务。ORCID 目前提供的学术出版物主要是文章、基金、专利等,更多学者的科研信息尚未包含其中。但 ORCID 旨在收集所有与作者相关的学术记录,从研究数据集、程序代码、博客文章到各种学术贡献。为实现这一功能,ORCID 将与 CrossRef、DataCite,以及各种数据服务商保持着密切合作。

#### 4.3 两种解决方案比较

个人名称统一标目的建立原则是通过附加内容对名称标目进行区分,这使得数据库中的统一标目项不够简洁,并且区分度降低,而且极易造成内容著录混乱,比较突出的是对学科、职业等附加内容的著录。在标目内容上,信息组织所追



求的是标目成分的准确性和稳定性,而用户检索则希望名称标目通俗、便捷,这二者之间本身就很难权衡。在标目形式上,个人名称规范数据库主要通过字典式的数据表对作者名称各种标识信息进行静态列举式关联,缺乏语义层面的描述和规范<sup>[17]</sup>。此外,由于各国在语言、文化背景等方面存在诸多差异,要想实现全球个人名称统一的标目形式是不现实的,只能通过技术手段来实现不同规范数据的集成与共享<sup>[18]</sup>。这些是通过构建个人名称规范数据库解决学者身份识别问题的局限所在。

不同于图书馆传统规范控制方法,以国际知名出版集团为代表的组织机构采用了唯一标识符的理论来解决学者身份识别问题。这一解决方案引起了全球出版社、平台商和学术团体等机构的极大关注,并得到更多的社会认同。然而现有学者身份唯一标识符系统类型过多,并且编码不规范,有许多待改进之处。唯一标识的理念在许多领域和行业都有应用,典型的例子有身份证号、银行账号、车牌号、产品条形码和电话号码等,图书期刊领域的 ISSN、ISBN、ISRC、CODEN,以及万维网的 URL、URI 等编码。学者身份唯一标识符的研究最近才得到较多关注,过去很多机构在标识符编码规范方面比较随意,缺少整体规划。目前最为流行的 ORCID 标识符有 16 位数字构成,但不包含任何语义信息,仅表示创建顺序。全世界图书总量远大于学者总量,ISBN 号使用 13 个数字就囊括了人类社会正式出版的所有图书,然而 ORCID 却使用了 16 位数字的编号系统,可见没有借鉴 ISBN 号的编码经验,实在有些遗憾。

## 5 学者身份识别的内在机制与关键技术

### 5.1 学者身份识别的内在机制

无论学者本名和异名在数据信息源中如何交叉重复,从知识组织的角度来看,学者名称歧义属于同义词和多义词问题。其中学者同名属于多义词范畴,需要处理的是一对多的关系,学者异名属于同义词范畴,需要处理的是多对一的关系。相比于主题概念间存在的同一关系、相关关

系和等级关系,学者名称只存在同一关系和相关关系,并且以同一关系为主,所以学者名称规范比主题概念规范易于实现。现有解决方案中,无论是以学者个人名称为主构建的统一标目,还是以字符编号为主构建的学者身份唯一标识符,其根本原理都是对学者个人名称进行规范控制。

在一定的数据信息源中,对学者个人名称建立具有区分度的统一标识以进行规范控制,是学者姓名识别的内在机制。根据上文所述学者同名出现的原理可知,学者名称统一标识符的区分能力与所在数据信息源的大小密切相关。简单而言,数据源集中的学者个体越多,越需设计区分表达能力强的学者身份统一标识符。图书馆多年实践证明,人名规范数据库中以学者名称为主构建的统一标目,其区分能力并不强,而且鉴于各国在语言、文化背景等方面的差异,要想实现全球人名统一的标目形式是不现实的。在最新版 RDA 规范中,虽明确定义了个人名称标识作为一种属性元素,但却并没有将其作为主标目,取代人名统一标目,在某种层面上仍延续了图书馆传统规范控制思想。

消除学者姓名歧义,进行学者身份识别,最终只能通过给定一个统一标识来解决。借助于编码符号表达法,设计具有足够大容纳能力和区分能力的学者身份唯一标识符系统,完全有可能实现。因此,为全球范围内所有学者设计唯一标识符,是消除学者名称歧义的最佳解决方案,并且一劳永逸,ORCID 等标识系统也正朝着这个方向努力。学者姓名统一标识符,作为学者身份的替代物,必须具有唯一性、稳定性和权威性,因此需要有固定的组织机构进行规范和管理。

### 5.2 学者身份识别的关键技术

学者身份统一标识符系统构建过程中涉及一系列复杂问题,诸如应为哪些学者设立统一标识符,如何有效编码,如何与学者各种异名进行关联,以及如何推广应用等。本文试着从解决这些问题的关键技术入手进行分析,以期为后续学者身份识别研究提供参考和借鉴。

#### 5.2.1 学者的识别与分析

图书馆构建人名规范数据库的原则是,只要

有其作品就为其做名称规范。由于馆藏图书并非全部都是学术著作,因此其人名规范数据库中的作者并非都是学者。Scopus ID、Research ID、ORCID 等标识符系统,主要采用系统分配和用户注册两种方式生成标识符,由于缺乏学者身份的确认和核查机制,其中网罗的个人有可能是学者之外的其他人员,如学生。严格来说,硕士生、博士生等不能完全等同于学者。学者应是具有一定的专业技能、知识水平、创造能力,在相关领域持续进行研究的人员。学生在学习阶段虽然有科技产出,但有些人毕业后可能会从事其他行业的工作,不一定继续进行学术研究。因此,在书目数据、期刊论文数据库等各种数据源中,如何识别与确认作者的学者身份,非常重要。一般来说,对具有专业技术职称的学者,尤其是具有教授、研究员等高级专业技术职称的学者,应该首先进行学者身份统一编号,并在编号的语义化表达设计中有所考虑。但对于还没有高级专业技术职称的其他专职研究者,是否需要结合发文量,以及如何结合以发文数量来确认其学者身份是值得探讨的地方。

### 5.2.2 标识符的语义化表达

据 Scopus 统计,全球大约有 2700 万研究者,包括政府和私营部门的研究机构,不包括学生<sup>[9]</sup>。对如此庞大的人员进行有效识别并非易事。符号编码主要有两种方式,一是纯数字式编码,二是数字与字母等相结合的混合编码。然而对数量庞大的实体进行编码是一项复杂而又系统的工作,稍不注意就会出现重号或者重复编号。最佳解决方法就是编码前对实体进行简单分类,并且将类别信息融入实体编码中,以增加编码本身的可辨识度。因此,一套成功的编码系统不仅需要考虑编码的稳定性、唯一性和权威性,还需要尽可能包含语义信息。

ISBN 号、个人身份证号码等号码系统都含有语义信息。比如 ISBN 号包含图书的国别、所在出版社、图书出版序号等信息,又如我国身份证号码中包含个人的出生时间、所在省份、市区等信息。尤为值得一提的是,科技界已认可的数字对象的永久性唯一标识符 DOI(Digital Object

Identifier, DOI)。DOI 作为数字内容的标识符,它不仅是一串复杂的字符编码,而是包含了与数字对象内容有关的语义信息的符号表达。DOI 标识符主要由前缀和后缀两部分组成,中间由斜线分隔。前缀是由 DOI 管理机构分配给 DOI 注册者(出版社)的号码,后缀是由已注册过的出版社对其数字内容进行组织分配的号码。后缀语法复杂,要求反映注册数字内容的层级信息,是一个连续的逻辑单元系统,并且可扩展。例如 Elsevier 的 DOI 标识符格式为:“DOI: 10.1016/刊物类型.期刊代码.出版年.出版月.论文流水号”。

目前以 ORCID 为主的已有学者身份唯一标识符,编码设计过于简单,鲜有包含与学者身份相关的语义信息。学者身份唯一标识符应充分借鉴 ISBN 号、个人身份证号码、DOI 等编号系统的经验,尽可能增添语义信息,可优先考虑增加较为稳定的属性信息,例如出生年月、出生地、性别、国别、民族等。一般而言,这些信息在用户提交注册表单时可自动获取,系统直接统一编号即可。

### 5.2.3 标识符与学者各种名称的关联

学者身份唯一标识符是消除学者名称歧义的最佳解决方案,但学者身份唯一标识符仅是学者身份的一种识别符号,它无法代替学者名称在现实社会中的意义。日常生活交流中,往往使用简短的姓名称呼对方,而不会使用特定身份证编号来称呼对方。在正式和非正式的学术交流中也一样,学者同样习惯使用简单的名称来署名各种学术记录,除非期刊出版有特别规定,否则不会使用一长串复杂的学术身份识别号。因此,学者名称歧义的消除最终需要将这唯一标识符与学者各种名称进行关联,才具有现实意义。

人名消歧技术是学者身份统一识别号与学者各种名称进行关联的基础。目前 ORCID 系统采用自动人名消歧和真实性认证相结合的方式。自动人名消歧主要根据作者的机构、地址、学科领域、文献题名、引文时间和合作者等信息进行识别,将同一个学者的不同名称划分成一个类别,然后分配统一的学者身份识别号<sup>[20]</sup>。然而自动人名消歧算法匹配结果的准确率无法达到 100%,

还需结合各种真实性认证方式,将学者身份统一识别号与学者各种名称进行关联。真实性认证主要包含自我认证和机构认证两种方式。自我认证是学者本人对自己的信息进行确认,由于系统无法保证用户都去进行确认,所以这种方式认证结果虽然最准确,但却有局限性。机构认证是由相关机构出面有专人负责学者个人信息的准确性进行确认。

其实,图书馆在构建人名规范数据库时进行了大量细致的学者姓名识别和认证工作。如果将已有中外人名规范数据库与学者身份唯一标识符系统进行对接,可大大提高人名消歧的效率,而且可确保学者身份统一识别号与学者各种名称关联的准确性。但值得注意的是,人名规范数据库主要以出版过图书的学者为处理对象,而学者身份唯一标识符系统则主要以发表过论文的学者为处理对象,这两个学者群体是交叉的,不完全一致。在具体处理过程中,需进一步甄别。

#### 5.2.4 标识符的推广和应用

人名规范数据库起初由图书馆业务部门负责单独建库,建成后再与书目数据库进行挂接,实施控制。由于人名规范数据库一直用于图书馆内部书目数据的组织与检索,是一个封闭的系统,严重制约了人名规范数据库的对外发展。以 ResearcherID 和 ORCID 为代表的学者身份唯一标识符系统,采取了与图书馆完全不同的方法,它们是一个完全开放的系统,允许学者个人免费注册。截至 2014 年 1 月已有 38.8 万名学者注册申请 ResearcherID。而 ORCID 从创始第一天起,就将包含 ResearcherID, AuthorID 等在内的其他作者标识符系统纳入其发展计划。目前 ORCID 已与 ResearcherID 完成了深度融合,接下来将全面推进与 Scopus、ACM 等数据库的整合工作。

学者身份唯一标识符系统能否成功地推广和应用,主要取决于该号码能否被整合到各个利益相关者的业务系统中,真正起到关联各种信息,成为数据交换点的作用。正因如此,学者身份唯一标识符绝不是一串孤立的编码符号,它不仅需

关联学者的背景和真实信息,而且需关联学者所有的学术记录,这在其推广和应用过程中至关重要。成功的学者身份唯一标识系统,在设计之初就需要考虑各种生态系统建设。例如,学者从稿件提交、同行评议,直到成果出版,与出版商的互动过程中,可利用学者身份唯一标识符关联学者的研究背景和其他真实信息,以验证作者和预期评审专家之间的联系,从而改进出版社的决策。再如,学者某一成果出版后,可将学者身份唯一标识符连同成果元数据、个人概要等信息一起关联到各种出版商、平台商和数据服务商系统中,与学者的其他学术成果产生关联,最大范围内整合学者的所有学术记录。

## 6 结语

社会学极为关注符号支配的各种策略和机制,并融汇于中,环环相扣<sup>[21]</sup>。未来在线科研环境中,如果每本书籍通过 ISBN 号进行识别,图表、数据、文档、论文等各种形式的学术记录用 DOI 号进行标识,而研究人员则可以用其独一无二的 ORCID 号来证明身份,那么无论是以人为中心的知识组织,还是以内容为中心的知识组织,都将更多借鉴社会学领域的有关研究成果,将研究重点转向以符号支配的各种组织策略和关联机制的探索中。在知识组织层面上,ORCID、ISBN、DOI 等各种统一表达符号的功能是相通的,即为了解决事物的同名冲突和一物多名问题。而关联数据作为最新的知识组织方法,它可借助 URIs 标识整合事物的各种符号表达系统,为包含各种数据、信息和知识在内的细粒度化语义单元,生成互联网上的唯一标识 URIs,这不但解决了开放世界环境下的非唯一性命名问题,而且建立了更高层次的虚拟秩序。各种类型事物或资源转换为建立了统一标识符的细粒度语义单元后,它们内在关联的特征、结构和基本规律将是本文今后重点研究的内容。

## 参考文献

- [1] 学者[EB/OL].[2015-03-05].<http://baike.baidu.com/>.



- [2] 赵宇星, 马莉蓉, 冯筱. 中文个人名称规范标识的选择与著录[J]. 情报探索, 2008 ( 8 ) : 115- 117.
- [3][5] 一个改造科研生态系统的愿景[EB/OL]. [2015- 03- 05]. <http://www.editage.cn/insights/a-vision-to-transform-the-research-ecosystem>.
- [4] 刘炜, 张春景. 试论网络资源的规范控制[J]. 现代图书馆情报技术, 2008 ( 12 ) : 27- 31.
- [6] 龚天芳, 张成昱, 张蓓, 等. ResearcherID 现状分析及应用启发[J]. 图书情报工作, 2013 ( 2 ) : 40- 45.
- [7] 邱均平, 余厚强. 替代计量学的提出过程与研究进展[J]. 图书情报工作, 2013 ( 10 ) : 5- 12.
- [8] 崔春, 毕强. 虚拟国际规范文档 ( VIAF ) 项目进展[J]. 图书情报工作, 2014 ( 3 ) : 129- 134.
- [9][12] 曹玉强. 国家图书馆中文名称规范的探讨[J]. 图书馆建设, 2007 ( 3 ) : 46- 48.
- [10] 卜书庆. 国家图书馆中文书目规范控制现状及研究[J]. 图书馆论坛, 2010 ( 10 ) : 209- 213.
- [11] 规范控制[EB/OL]. [2015- 03- 05]. <http://zh.wikipedia.org/wiki/%E6%AC%8A%E5%A8%81%E6%8E%A7%E5%88%B6>.
- [13] 朱青青, 孙凤玲. FRAD 与中文名称规范控制之研究[J]. 国家图书馆学刊, 2012 ( 2 ) : 19- 22.
- [14] RDA 发展联合指导委员会. 资源描述与检索 ( RDA ) [S]. 国家图书馆出版社, 2014 ( 4 ) : 451- 499.
- [15] 陈金星, 祝忠明. 责任者名称规范控制研究及进展[J]. 现代图书馆情报技术, 2009 ( 12 ) : 12- 17.
- [16][17] 孙红, 万邵华. ORCID: 数字化背景下专业人员的学术身份证[J]. 长江大学学报 ( 社科版 ), 2014 ( 8 ) : 191- 192.
- [18] 刘小玲. 关注国际编目进展强化书目规范控制[J]. 图书馆论坛, 2008 ( 2 ) : 86- 89.
- [19] 刘润达, 王运红. 开放研究人员及贡献者唯一标识 ( ORCID ) 概述[J]. 情报科学, 2013 ( 11 ) : 86- 90.
- [20] Scopus Author Identifier[EB/OL]. [2015- 03- 05]. [http://help.scopus.com/Content/h\\_authr\\_intro.htm](http://help.scopus.com/Content/h_authr_intro.htm).
- [21] 宫留记. 是哲学家还是社会学家——从布迪厄的学术轨迹评其学术身份[J]. 广西社会科学, 2007 ( 3 ) : 27- 30.

作者简介 常娥 ( 1979- ), 女, 博士, 东南大学图书馆副研究馆员。

收稿日期 2015- 04- 16

( 上接第 124 页 )

- [50] Landoni, M., and P. Diaz. E- education: design and evaluation for teaching and learning[EB/OL]. [2015- 02- 06]. <http://journals.tdl.org/jodi/index.php/jodi/article/view/118/65>.
- [51] Shelburne, W. E- book usage in an academic library: user attitudes and behaviors[J]. Library Collections, Acquisitions and Technical Services, 2009, 33 ( 2/3 ) : 59- 72.
- [52] Gibson, C., and F. Gibb. An evaluation of second generation e- book readers[J]. The Electronic Library, 2011, 29 ( 3 ) : 303- 319.
- [53] Rozell, E., W. Gardner. Cognitive, Motivation and Affective Processes Associated with Computer- related Performance: a Path Analysis[J]. Computers in Human behavior, 2000, 16: 192- 222.
- [54] Letchumanan, M., R. Tarmizi. E- book utilization among mathematics students of University Putra Malaysia [J]. Library Hi Tech, 2011, 29 ( 1 ) : 109- 121.
- [55] Biranvand, A., A. Khassheh. E- book Reading and its Impact on Academic Status of Students at Payame Noor University, Iran[EB/OL]. [2015- 02- 06]. <http://digitalcommons.unl.edu/libphilprac/1170>.
- [56] Wilson C. J, D'Ambra and R. Drunnon. Exploring the fit of e- books to the needs of medical academics in Australia[J]. Electronic Library, 2014, 32 ( 3 ) : 403 - 422.
- [57] Hwang, J, J Kim, J Lee, et al. Usage patterns and perception toward e- books, experiences from academic libraries in South Korea[J]. Electronic Library, 2014, 32 ( 4 ) : 522- 541.
- [58] Lapa C. The normal distribution in LibraryThing: investigation on the combined literature of fifteen books[J]. Collection Building, 2013, 32 ( 4 ) : 145- 153.

作者简介 赵珊珊 ( 1978- ), 女, 硕士, 上海对外经贸大学图书馆馆员。

收稿日期 2015- 04- 08