

基于ISNI的学术应用生态构建^{*}

刘翔¹ 黄晨²

(1. 浙江理工大学信息化办公室, 杭州 310018; 2. 浙江大学图书馆, 杭州 310027)

摘要: 中国工程科技知识中心项目建设中涉及大量学者、机构的标识, 为有效解决“谁是谁”的问题, 采用国际标准名称标识符 (ISNI) 解决名称的唯一标识。在此基础上进一步扩展到跨机构的应用, 力求通过生态圈的建设服务于多行业, 进而形成学术应用的生态。

关键词: 国际标准名称标识符; 统一身份认证; 机构知识库

中图分类号: G205

DOI: 10.3772/j.issn.1673-2286.2020.05.008

1 ISNI的产生背景

学术研究的主体是科研机构与科研专家, 对科研机构和科研专家的信息进行全面、准确地收集、描述和关联, 进而形成丰富、鲜活的机构专家知识库, 是知识时代重要的信息基础设施, 也是一项烦杂而艰巨的工作。国内外各类机构对机构专家库从理论和实践方面开展了诸多研究, 成绩斐然。谷歌学术、微软学术以及百度学术等互联网公司凭借其强大的信息收集和处理能力, 建设了一批大而全的机构、学者数据库^[1-2], 而由康奈尔大学启动的VIVO项目, 是一个包含人员、机构信息的本体库, 能够实现机构、人员、学术成果的互联^[3]。这些知识库的构建过程中, 名称消歧是最基础和核心的工作, 为此以美国国会图书馆等机构建设的虚拟国际规范文档 (VIAF) 项目^[4]为代表, 基于知识积累的名称规范工具受到业界关注, 并广泛应用。各种基于规则、机器学习等算法的消歧、名称归一化研究也广泛开展, 这些都可以归结为名称唯一性问题。尽管有了以上研究和成果, 但仅通过名称字符串规范和字符串纠偏等方式的解决途径, 要实现对名称唯一性问题的较好解决, 仍然任重道远。为此, 相关机构从源头着手, 采取对命名实体 (人、机构、成果等) 赋予唯一值的方式 (ID模式), 以期从根本上解决问题。出版社和数据库商发起

的Orcid^[5]、Clarivate Analytics的ResearcherID^[6]、国际数字对象识别号基金会的DOI^[7]就是这种解决方案的实践代表, 并已经取得成功。在名称唯一性问题的标准化工作方面, 国际标准化组织 (ISO) 在2012年发布了国际标准名称标识符标准 (International Standard Name Identifier, ISNI), 即ISO 27729:2012^[8]; 我国在2012年由原新闻出版总署发布了名为中国标准名称标识符的行业标准, 规范了新闻出版行业纸质出版的作者标识^[9]。此外, 中国科学院文献情报中心和ORCID合作, 提出了iAuthor名称标识符系统^[10]。

从上述研究和实践, 可以看到我国在人、机构名称知识库建设中采用ID方式作为名称唯一性问题解决方案的大规模、有影响力的项目不多。中国工程院国际工程科技知识中心 (以下简称“知识中心”)^[11]建设的中国工程科技研究机构与专家库 (以下简称“知识中心机构专家库”) 正是我国为数不多, 且较有影响的专业知识库, 该知识库对中国工程科技界的专家学者进行全面学术揭示, 相较于谷歌学术、微软学术以及百度学术建设的大而全的学者数据, 该知识库定位更加精准和高端。中国工程科技机构与专家库项目于2014年立项, 至今清洗梳理了955万条专家数据, 包含8 000多万条中文学术论文信息、350万条SCI引文数据、80多万条EI工程索引、1 100万条专利数据, 近300万条专著和80多万

^{*}本研究得到中国工程科技知识中心项目“中国工程科技研究机构与专家库” (编号: CKCEST-2020-1-13) 资助。

条成果数据^[12],已经成为我国较全面的机构专家数据库之一。知识中心机构专家库建设过程中遇到的最大问题是对专家姓名的唯一性认定,就是要解决“谁是谁”的问题。如何构建专家唯一标识符系统,尽管有较多的国内外做法可以复用,但这往往严重依赖于境外服务器和国外标准及知识产权,对于我国的学术信息安全存在一定程度的隐患;同时,随着互联网的飞速发展,各个行业对作者标识的需求不断增加,如对于学者的网络漫游认证、各种信息平台互通等方面,都需要有国家标准来进行统一。在这个统一的标准下,对学者学术研究追踪的基础上,通过跨行业跨机构的应用生态的构建、关联,进而实现多元学术价值增值和知识共享。

2 ISNI发展现状与特征

2.1 ISNI简介

(1) ISNI标识符。ISNI的基本目标是提供一个持久的标识服务,其目标是识别跨多个创造性活动领域的公共身份,减少重复劳动并提高准确性,它是继ISBN、ISSN、ISRC和DOI等国际著名代码之后,为在数字环境下解决数字参与者身份唯一标识诞生的国际性代码,以解决搜索和发现中的名称模糊问题,追踪确认识识创造者的成果。

ISNI的编码由16位数字组成,包含两个部分,即15位十进制数字和一位校验位,一个编码只包含数字,不包含任何含义,其中的任何部分不传递信息,可分配给研究人员、发明人、作家、艺术家、视觉创作者、表演者、制作人、出版商和研究机构等知识创造者。编码在书写、印刷或其他使用场合,以可读的形式呈现,如ISNI 1422 3422 5434 7563。

(2) ISNI编码分配、解析与数据更新。ISNI编码分配工作是通过其成员单位提供的基本数据,由法国国家图书馆和大英图书馆采用查重算法与ISNI数据库中已有的数据比对,分配新的标识符。采用“[http://isni.org/isni/\[isni\]](http://isni.org/isni/[isni])”以关联数据的方式进行解析,提供html和XML格式的ISNI核心元数据,详细展现知识创造者或者机构的信息,以使用户进行身份确认。对于已分配的编码,可通过网页表单或Atom PubAPI的方式进行数据更正或更新。

(3) ISNI发展现状与其他标识符关系。目前,ISNI已经分配了超过1 102万个标识符,其中包含291万

名研究者和93万家机构,全球已有OCLC、哈佛大学、Youtube、VIAF等54个大型机构与ISNI展开了合作。众所周知的ORCID也与ISNI签署了合作协议备忘录,成为ISNI的重要成员,并在2013年通过RingGold加入ISNI,成为ISNI的代理注册机构,在学术研究领域进行标识符分配^[13]。

2.2 ISNI标准

鉴于ISNI的国际权威性,知识中心经过与ISNI组织协商,成为ISNI在中国的唯一代理,在此基础上起草了《中国标准名称标识符》标准草案,目前已经通过全国信息与文献标准化技术委员会的审定。

由于ISNI不支持标识符的预分配,知识中心构建名为KnowledgeID的学者唯一标识,简称KID。KID的编码由10位不包含任何意义的数字组成,由系统随机分配,知识中心对专家库项目构建的900多万名专家进行了唯一标识符的分配,利用KID进行解析,以定位任意专家的学术信息^[14],用于国内的学者识别。对于在系统中已经认领本人成果的学者,系统将有代表性成果同步到ISNI组织用于分配ISNI标识符,并将KID与ISNI的标识符进行绑定,同时在KID系统提供无缝解析。类似互联网的CN域名,KID系统可以确保中国学者解析服务的持久安全运行。

截至2019年,基于KID的桥梁作用,为知识中心同步分配了总计21万名中国专家的ISNI标识符和相应的解析服务,并应用到知识中心的用户认证系统。在《中国标准名称标识符》标准草案获得采用后,将对KID系统从面向知识中心的服务,提升为服务于全国各领域、各行业的需求,进行系统功能的扩展和标准化,真正成为ISNI在中国的标识符分配和解析系统。

3 ISNI的应用生态建设

单独的学者标识符、专家库作为特定机构和领域的需求无法形成应用共同体,必须经过上下游结合才能形成良性循环的生态。ISNI标准的唯一性、公开性、国际性和隐私安全性,既能保护使用过程中的隐私,又可鉴别参与者的学术信用和学术水平。学者个人、学术机构、政府组织都可以因为ISNI标识符的识别作用而获益。学者因为唯一标识符的使用可以自动收集自己的学术信息,构建个人学术成果库;机构可以利用唯一

标识符确定学者的学术画像,构建机构知识库,研判评价学术趋势;政府组织利用唯一标识符获得学者或机构的学术征信信息,用于基金、项目的评审和委托,学科评价、人才引进等各个方面。随着用户数据的沉淀,基于数据驱动的信任机制将不断健全,在此基础上可提供的服务内容也必然更加丰富,吸引用户更广泛地参与。作为基础标准,ISNI为建设开放合作的跨机构、跨行业的应用奠定了基础。

目前,通过开放基于KID系统的学者认证入口,打通学者、学术机构和学术发布平台(期刊),初步形成一个学术应用生态(见图1)。



图1 基于ISNI的应用生态

3.1 服务学者与机构

目前,中国工程科技机构与专家库全面启用了ISNI标识符,对于由专家库系统分配的ISNI标识符,可以在ISNI网站进行解析,并且获得详细的链接信息,直接指向中国工程科技机构与专家库页面,为学者提供持久的KID解析服务,如图2所示。

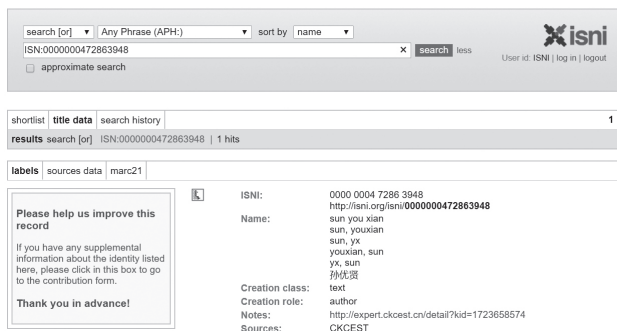


图2 ISNI解析页面

基于KID解析服务,研发了专家名片小程序服务于学者的学术网络社交。利用专家名片小程序学者可以快捷地创建个人学术名片,由于数据源于中国工程科技机构与专家库,用户在创建名片后,立即得到个人成果数据的推送,便于利用碎片时间完成个人学术成果的确

认,快速构建自己的云端成果库。有别于传统的商务名片应用,学者名片主要是作为学者的学术形象构建。因此,学者通过微信小程序进行名片社交分发的过程中,接收者可以直接在名片上面看到学者的学术画像,从而详尽地了解学者的学术背景,如图3所示。



名片创建与更改

名片导出与分享



名片收藏与通信

专家成果与解析

图3 专家名片

3.2 打通互联网应用

基于KID解析服务,建立学者唯一标识符的解析系统,利用这个解析系统可以实现学者身份的认定,打通互联网应用。目前,已经完成基于OAuth2.0协议和Shibboleth协议的学者唯一标识符身份认证服务,并集

成到大学数字图书馆国际合作计划(CADAL)的门户网站上为其提供服务。

用户在CADAL门户的登录过程中,除传统的用户名密码认证之外,同时提供微信“扫一扫”的服务,这一服务正是基于专家名片小程序的功能来完成的。新用户首次使用“扫一扫”功能登录时,会被引导到专家名片小程序生成名片。用户在注册的同时生成自己的学者名片,既成为CADAL的用户,也成为专家库的用户。注册用户通过“扫一扫”就可以方便地进入CADAL门户,享受相应权限的服务。目前,合作厂商将这一身份认证模式接入了大连海事大学等16所高校的统一身份认证系统,用户只需要通过简单的“扫一扫”就可以完成身份识别,进入学校网络以获取相应权限的服务^[15]。

基于这一服务的便捷性,与中国教育科研网CERNET-CARSI达成合作,校园网用户有望通过简单的“扫一扫”实现校园网际的漫游。更加值得期待的是,通过这一身份认证方式可以替代传统的VPN服务。一个合法的用户,只要通过自己的手机扫码,就可以在任何地方获得校园网内的资源服务以及其他的内部行政服务。

3.3 助力出版与知识产权

目前,同多家期刊出版社已就ISNI在出版行业的应用进行了较为深入的研究,并认为期刊出版社采用ISNI标准,为出版和知识保护提供了方法。同时,中国版权保护中心也在密切跟进ISNI在知识产权保护方面的应用。中国工程院院办期刊*Engineering*杂志已经率先启用KID的解析服务,编辑部要求投稿者必须注册获得学者唯一标识符,并且在稿件上注明自己的唯一标识符。这样编辑可以方便地通过唯一标识符解析系统获取投稿者的学术画像,了解投稿者的学术背景,以便更准确地评估稿件内容,录用稿件将标记上学者的ISNI唯一标识符。

3.4 服务学术社交网络

在国外学术社交网络ResearchGate、Academia.edu和Mendeley等兴起的浪潮下,国内也产生了一批致力于学术交流和科研服务的在线平台,如科学网、小木虫和百度学术等。尽管现有的学术社交网络如雨后春笋般涌现,但是这些网络平台在服务科研活动的专业性上一直受到用户的广泛质疑,如学术社交网络中充斥

着大量与学术无关的内容,学术信息质量得不到保证、“学术功能”有余而“社交功能”不足、学术资源组织方式落后和论文知识版权纠纷等诸多问题。ISNI标识数字知识的属性,为从根本上推动学术网络社交提供了解决方案,具体表现在:①促进跨平台互通,根据学者ISNI标识符,打通学术社交网络的边界,实现国内外学术社交网络的互联互通和学术交流互信;②促进学术社交内容专业化,学者的ISNI标识符充分揭示其研究领域、研究方向和研究成果等,能提高学术信任度,解决学术网络社交的虚拟身份造成的学术不规范行为;③促进学术交流,ISNI标识符揭示学者的学术画像,可促进学者学术交流;④保护知识创新,学术社交网络会产生创新的知识单元,ISNI标记数字知识的属性,规范了网络虚拟学术交流。

4 结语

基于ISNI的学者唯一标识符系统应用前景十分广泛,除能够准确确定学者的学术信息,有利于完成学者的学术画像,准确定位学者的学术生命周期外,还可以作为学者的唯一身份识别码,在各个学术机构学术应用场景中实现无缝漫游,成为建立学术应用生态圈的基础。

参考文献

- [1] Microsoft Academic [EB/OL]. [2020-04-23]. <https://academic.microsoft.com/home>.
- [2] 学者主页 [EB/OL]. [2020-04-23]. <http://xueshu.baidu.com/usercenter/data/authorchannel?cmd=frontpage>.
- [3] VIVO Technical Documentation Archive [EB/OL]. [2020-04-23]. <https://wiki.lyrasis.org/display/VTDA/VIVO+Technical+Documentation+Archive>.
- [4] VIAF: The Virtual International Authority File [EB/OL]. [2020-04-23]. <http://www.viaf.org>.
- [5] ORCID Connecting research and researchers [EB/OL]. [2020-04-23]. <https://orcid.org>.
- [6] Welcome to the new Web of Science ResearcherID [EB/OL]. [2020-04-23]. www.researcherid.com.
- [7] Frequently Asked Questions about the DOI System [EB/OL]. [2020-04-23]. <https://www.doi.org>.
- [8] ISO Publishes the ISNI Standard (ISO 27729:2012) [EB/OL].

- [2020-01-10] . <http://www.isni.org/content/iso-publishes-isni-standard-iso-277292012>.
- [9] 赵树旺. 中国数字出版内容国际传播研究 [M]. 北京: 中国传媒大学出版社, 2016.
- [10] 张建勇, 黄永文, 于倩倩, 等. 中国ORCID注册平台iAuthor的设计与实现 [J]. 现代图书情报技术, 2015 (3): 84-91.
- [11] 国际工程科技知识中心 [EB/OL]. [2019-12-08]. <http://www.ikceest.org/index.htm>.
- [12] 中国工程科技研究机构与专家库 [EB/OL]. [2019-12-20]. <http://expert.ckceest.cn>.
- [13] ORCID to begin using Ringgold as Registration Agency [EB/OL]. [2020-01-10]. <http://www.isni.org/content/orcid-begin-using-ringgold-registration-agency>.
- [14] KID解析注册平台 [EB/OL]. [2020-01-10]. <http://kid.ckceest.cn>.
- [15] 大连海事大学统一身份认证系统 [EB/OL]. [2020-01-12]. <https://hall.dlmu.edu.cn>.

作者简介

刘翔, 男, 1977年生, 硕士, 副研究馆员, 研究方向: 数字图书馆及数据分析。

黄晨, 男, 1970年生, 硕士, 研究馆员, 通信作者, 研究方向: 数字图书馆与大数据应用, E-mail: huangc@zju.edu.cn。

Construction of Academic Application Ecology Based on ISNI

LIU Xiang¹ HUANG Chen²

(1. Information Office Zhejiang Sci-Tech University, Hangzhou 310018, China; 2. Library of Zhejiang University, Hangzhou 310027, China)

Abstract: The identification of a large number of scholars and institutions is involved in the construction of the China Knowledge Centre for Engineering Sciences and Technology, in order to effectively solve the problem of "who is who", the research group chose the ISNI to solve the unique identification of names. On this basis, it is further extended to the application of cross institutions, striving to serve multiple industries through the construction of ecosystem, and then form the construction of academic application ecology.

Keywords: ISNI; Uniform Identity Authentication; Institutional Repository

(收稿日期: 2020-02-07)