

知识组织领域标准规范发展研究

曹树金 曾盈盈 廖赛源 曹茹烨

摘 要 在信息泛化、新技术不断涌现的环境中,推进知识组织领域的标准化具有重要意义。本文收集了2009年以来ISO、IFLA、W3C以及美国、英国和中国的知识组织领域标准规范,从类别、发布时间、标准状态与标龄、引用关系四个方面进行描述和比较,全面呈现知识组织领域标准规范发展状况。知识组织领域标准规范呈现资源描述标准更新、标识标准被全面修订、大数据标准化工作取得进展等重要变化,以及体系日益完善、元数据标准等部分标准发展良好、中国标准对国际标准的同步性和引领性有待加强等特点,同时知识组织领域的标准规范逐渐呈现出交叉融合、建设机制更加灵活、知识关联与数据交换逐渐成为重点等发展趋势。建议我国既要根据知识组织领域的发展需求对传统标准规范进行不断更新,也要探索新技术在知识组织领域的应用标准;既要与国际标准保持同步,又要力争发挥引领作用。图10。表2。参考文献18。

关键词 知识组织 标准规范 标准化组织 图书情报学

分类号 G254

The Development of Standards in the Field of Knowledge Organization

CAO Shujin, ZENG Yingying, LIAO Saiyuan & CAO Ruye

ABSTRACT

The development of the Internet, big data and intelligent technology leads to significant changes of knowledge organization, and the corresponding theory, method, technology and tool also have changed. As new technologies appear, knowledge organization shows the characteristics of multi-granularity and fine granularity. At the same time, the focus is shifting: the connection between knowledge and the construction of the network is paid more attention. Therefore, in order to avoid unnecessary exploration in practice, and facilitate the management and interactive operation of systems in the field of knowledge organization, it is of great significance to promote the standardization in the field of knowledge organization.

Therefore, this study mainly explores whether the standards and norms in the field of knowledge organization have been formulated and updated in a timely manner in the new information environment, and the characteristics and development trend of current standards and norms. This study sorts out the ISO, IFLA, W3C, ANSI, BSI and Chinese knowledge organization standards in the past 13 years, then describes and compares these standards from the aspects of quantity, release time, standard status and age, citation relationship. Important changes have been made in the field of knowledge organization, such as updating the

通信作者:曹树金,Email:caosj@mail.sysu.edu.cn,ORCID:0000-0003-1855-4522(Correspondence should be addressed to CAO Shujin,Email:caosj@mail.sysu.edu.cn,ORCID:0000-0003-1855-4522)

resource description standards, revising the identification standards comprehensively, and making progress in big data standardization. The knowledge organization standards system is improving day by day. Some standards such as metadata standards are well developed, and the synchronization and guidance of Chinese standards to international standards need to be strengthened.

It is found that the standards show trends of cross-integration, more flexible construction mechanism and gradually shift of focus on knowledge association and data exchange. At last, it is suggested that we should not only continue to update traditional standards in accordance with the development of the knowledge organization field, but also explore to apply emerging technologies into knowledge organization standards. At the same time, we should also keep pace with international standards and strive to play a leading role. 10 figs. 2 tabs. 18 refs.

KEY WORDS

Knowledge organization. Standards. Standardization organization. Library and information science.

0 引言

近年来,互联网的进一步渗透、信息的泛化、大数据的涌现及智能技术的发展和應用都给知识组织环境带来了较大的变化。关联数据、语义网、本体和知识图谱的快速发展,使得知识组织的理论、方法、技术、工具等发生变化。知识组织呈现出新的特征,如组织单元呈多粒度化,细粒度到了实体;更加注重知识之间的关联,知识网络具有更复杂的关系;由计算机借助智能技术开展知识组织的情况逐渐增多,等等。因此,推进知识组织领域的标准化具有重要意义,一方面可以避免实践过程中不必要的重复探索,有效提高知识调用的效率;另一方面不同知识组织体系的构建、管理、应用可以遵循共同的标准,使知识组织领域中的各体系更便于互操作和数据交换。

在知识组织领域标准化方面,国际组织及各国都做出过积极的探索,编制并发布了一系列标准和规范。但是新的信息环境下知识组织领域的标准规范是否得到了及时制定和更新,现行标准规范呈现出哪些特点,未来发展趋势如何,这些问题都有待回答。考虑到2009年是新世纪以来我国知识组织领域相关标准发布的第一个峰值年,因此,本文以2009年为研究开

端,对国内外近13年来知识组织领域标准规范进行梳理,分析发展现状与趋势,以便反映知识组织发展的方向;同时通过对比国内外标准制定的情况,为我国相关标准的制订与修订提供参考。

1 研究综述

国内外关于知识组织领域标准规范的研究,主要从两个视角出发。

一是从宏观方面分析知识组织或信息组织领域标准规范的发展情况。①梳理相关领域标准的发展情况,如分析数字环境中有关KOS开发和编码的国家和国际标准的最新进展^[1],归纳正式成文的知识组织体系标准以及知识组织体系的事实性标准的发展情况^[2],总结知识组织规范呈现出的多类型、多结构、多元化的发展趋势^[3]。②关注知识组织体系标准的特征与变化,如认为叙词表编制标准依旧是知识组织体系标准中的重心^[2],专业标准为主、通用标准为辅,标准起草机构逐渐多元化,标准更新维护起步较晚且时间集中,也存在信息组织标准更新频率有待加快、国际权威性有待提升等问题^[4]。③对国内外相关标准规范进行对比,如通过分析标准规范的内容^[5]和开发管理层面的发展情况,发现我国信息组织领域标准规范存在采标

率低、采标情况混乱,研究滞后、权威性差、整体标龄情况不容乐观等问题^[6]。相对来说,国内学者比国外学者更重视从宏观上分析知识组织领域标准规范的发展情况。

另一种是从微观方面讨论知识组织细分领域标准规范的编制、使用和发展等问题。在资源描述方面,国外学者对资源描述与检索标准(RDA)的原则和概念进行介绍^[7],论述了由国际图书馆协会联合会(IFLA)发布的《国际标准书目描述》(ISBD)的主要特点^[8];我国学者介绍了《资源描述》国家标准的起源和技术特点,从语义网信息组织的角度分析了领域标准规范的发展趋势,包括描述和组织的融合、规范和描述一体化、信息组织标准植根于网络技术标准中^[9]。在元数据方面,有研究介绍了用于社会科学信息组织的元数据标准DDI的内部结构、概念模型以及与其他元数据的兼容性问题^[10];通过对图书馆元数据标准的回顾,指出当前元数据标准是与语义网标准相当的数据组织的基础^[11]。此外,有研究关注到国际标准关联标识符(ISLI),分析ISLI在知识服务领域的应用价值,并提出ISLI可以实现异构知识资源的有序对齐和匹配^[12];探讨ISLI在图书馆行业的应用,并建议参照ISLI对已有出版物进行回溯关联编码,基于ISLI对知识元进行标注和管理^[13]。

综上所述,已有关于知识组织领域标准规范的研究并不多,尤其是缺少对近些年知识组织领域标准规范整体发展情况的分析。知识组织标准规范会随着技术发展和社会需求而不断变化,因此,本文拟通过对近年来知识组织领域标准规范开展系统分析,弥补现有研究的不足。

2 知识组织领域标准规范的定量分析

2.1 数据来源

许多研究者在探讨知识组织的相关问题时,往往无法完整准确地界定知识的范畴,从而难以清晰界定知识组织的理论基础和领域范畴^[14]。虽然在我国有观点认为知识组织是信息

组织的新发展,但是国内外,尤其是国际上,更多的是将“知识组织”和“信息组织”交替使用,在维基百科中“知识组织”为首选术语,而“信息组织”作为变体;这两个术语有时也会合并为“知识和信息组织”来使用^[15]。由此可见,有时并不能严格区分“知识组织”和“信息组织”,因此在本研究的标准选择中不对这两个概念加以严格区分。为避开概念界定中关键词选择和描述的问题,本文构建了知识组织标准分类框架,并在检索时使用分类框架中的具体类别名称进行检索,尽可能全面收集近年来知识组织领域的标准规范。

遵循权威性、全面性、即时性等原则,本文选取国际标准化组织(ISO)、国际重要行业组织(IFLA、W3C)以及美国 and 英国国家标准机构、中国国家及行业和地方标准机构发布的标准规范为数据源,因法国、德国、日本等国家知识组织领域标准数量较少,因此未纳入考量。根据国际标准分类法(International Classification for Standards, ICS)和中国标准分类法(Chinese Classification for Standards, CCS),人工收集和筛选出2009年之后颁布、更新和修订的知识组织领域的现行标准规范,共计179个。由于资源环境所限,所收录的标准仍有未尽之处。其中,ISO标准和美国、英国国家标准的选择以ICS类表中“01类(综合、术语学、标准化、文献)”和“35类(信息技术、办公机械)”为依据,共收集到43个ISO标准、21个美国国家标准和12个英国国家标准;所选IFLA标准规范来源主要包括“资源描述”“术语与词汇”“概念模型”几大板块,共收集到15个IFLA标准;W3C标准规范以“Standards”板块为来源,对大数据环境下元数据相关标准、关联数据、网络本体、数据交换等方面进行检索与筛选,共收集到13个W3C标准;中国标准以CCS中“A14图书馆、档案、文献与情报工作”类目下收录的标准为主要研究样本,同时收集“L00信息综合技术”类目下的相关标准,共筛选出41个国家标准和34个行业与地方标准。

2.2 标准类别、发布时间、标准状态与标龄分析

2.2.1 分类框架

本研究根据标准所规范的领域对象将知识组织领域标准分为术语和标识标准、类表和词表及其编制和互操作标准、分类和标引及索引标准、元数据及描述标准、关系网络构建标准五大类。如图1所示,术语和标识标准包括术语和

词汇标准、标识标准;类表和词表及其编制和互操作标准包括分类表和叙词表及其编制和互操作标准;分类和标引及索引标准包括分类和标引标准、索引标准;元数据及描述标准包括元数据标准、描述标准;关系网络构建标准包括本体及其互操作标准、主题图互操作标准、知识关联标准。

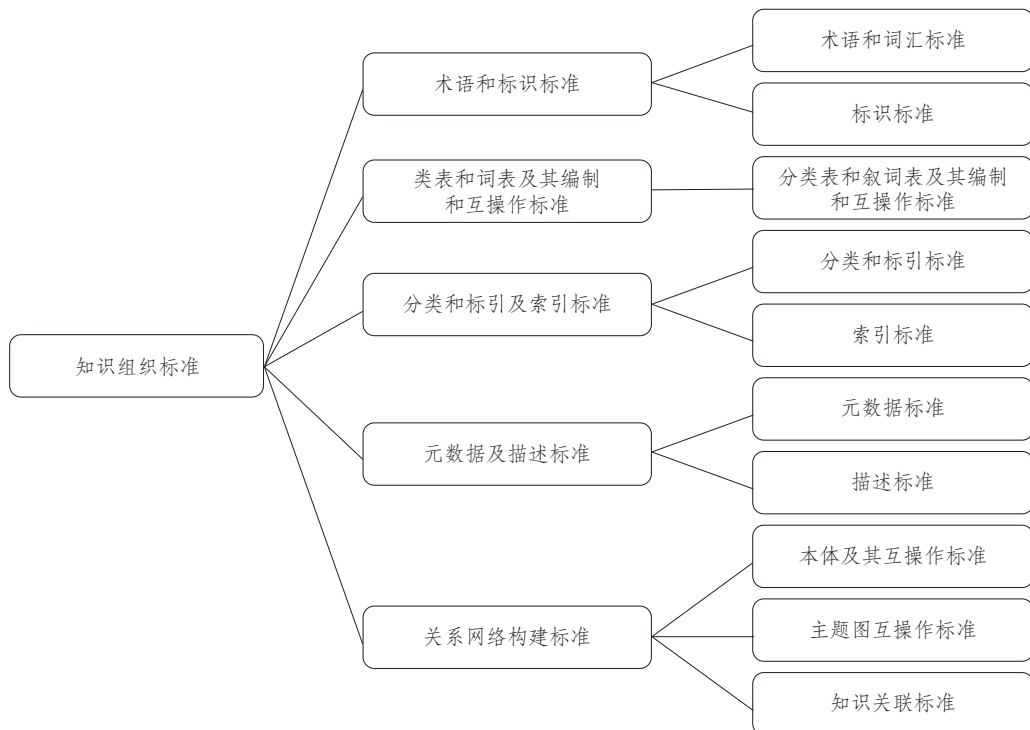


图1 知识组织标准分类框架

2.2.2 标准类别分布

本研究共收集到近13年来新发布和更新的知识组织相关标准151个(见表1)和28个知识组织相关规范。151个相关标准中术语和标识标准有42个,类表和词表及其编制和互操作标准有10个,分类和标引及索引标准有6个,元数据及描述标准有67个,关系网络构建标准有26个。从小类别来看,元数据标准最多,有62个;其次是标识标准,有36个;索引标准、主题图互操作标准及知识关联标准较少。151个标准中,ISO标准和中国国家标准较多,分别有43个和

41个,中国国家标准中有20个是采用的ISO标准;美国国家标准和英国国家标准相对较少,且英国国家标准大多采用的是ISO发布的标准。

从统计结果来看,作为国际权威的标准化组织,ISO发布标准数量最多,其中综合性标准居多,且涵盖范围广。中国国家标准化管理委员会发布标准数量与ISO发布数量相差不大,地方和行业标准数量有34个,说明近13年来我国对知识组织领域的标准化比较重视。

从各类别标准所涵盖的内容上看,术语和标识标准主要包含图书、档案等传统信息资源

表1 知识组织标准类别分布

类别	数量	标准	ISO 标准	中国标准			美国国家 标准	英国国家 标准	总计
				国家标准	行业标准	地方标准			
术语和标识标准	术语和词汇标准		4	2	—	—	—	—	6
	标识标准		10	13	8	—	2	3	36
类表和词表及其编制和互操作标准	分类表和叙词表及其编制和互操作标准		2	4	2	—	—	2	10
分类和标引及索引标准	分类和标引标准		—	3	—	—	—	1	4
	索引标准		—	1	—	—	1	—	2
元数据及描述标准	元数据标准		15	7	16	5	15	4	62
	描述标准		1	1	3	—	—	—	5
关系网络构建标准	本体及其互操作标准		11	9	—	—	3	1	24
	主题图互操作标准		—	—	—	—	—	1	1
	知识关联标准		—	1	—	—	—	—	1

和视听作品、网络数字资源等新型信息资源的基础术语、词汇及标识符,如 GB/T 34110—2017《信息与文献 文件管理体系 基础与术语》、WH/T 48—2012《信息与文献 数字对象唯一标识符系统》等;类表和词表及其编制和互操作标准主要用于规范多语种叙词表和主题分类词表,如 BS ISO 25964-1:2011《信息与文献—叙词表及与其他词表的互操作性—第1部分:信息检索用叙词表》等;分类和标引及索引标准集中于对各类资源建立分类与索引规则,如 GB/T 36070—2018《地方志索引编制规则》、GB/T 32153—2015《文献分类标引规则》等;元数据及描述标准方面,在科技人才元数据与学术规范元数据方面制定新标准,如 GB/T 35397—2017《科技人才元数据元素集》、WH/T 67—2014《期刊论文元数据规范》等,同时,对元数据交换等技术规范较为关注,发布了 INCITS/ISO/IEC 19503:2005《信息技术—XML 元数据交换(XMI)》等标准;关系网络构建标准梳理了本体、主题图、知识关联方面的标准,本体及其互

操作标准除对本体框架及不同领域资源的本体构建进行规范外,在文化遗产以及区块链技术方面也制定了新标准,如 ISO/TS 23258:2021《区块链和分布式账本技术—分类法和本体》、GB/T 37965—2019《信息与文献 文化遗产信息交换的参考本体》等,主题图与知识关联涉及标准较少,仅有 GB/T 38378—2019《新闻出版 知识服务 知识关联通用规则》和 BS ISO/IEC 13250-6:2010《信息技术—主题图—第6部分:压缩语法》。

另外,统计发现,IFLA、W3C 两机构分别发布 13 个和 15 个与知识组织相关的规范,规范主题可归纳为传统知识组织方法自动化和大众化技术与新兴知识组织技术两大类:传统知识组织方法自动化和大众化技术包括知识组织自动化技术(主要包括术语自动抽取、词表自动构建与丰富、自动标引、自动分类、自动映射)与用户参与式知识组织技术(大众分类法和网络百科);新兴知识组织技术包括语义网(如关联知识聚合)、本体、关联数据、主题图与知识图谱、

知识组织系统互操作与可视化技术,以及细粒度知识组织技术、大规模语义知识组织技术和视听资源知识组织技术等。

2.2.3 标准发布时间脉络分析

标准的发布时间是指标准经过立项、起草、征求意见、审查等一系列过程后正式发布的时间。为了理清不同类别标准的发展脉络,对分

类框架中每一类别标准的发布时间进行梳理(见图2)。总体来看,除部分发布数量较少的标准类别外(如分类和标引标准、索引标准、主题图互操作标准等),各类标准数量随时间推移呈波动上升趋势,2019年是13年来标准发布数量的峰值年,说明近年来,知识组织领域的标准制定正在逐步发展。

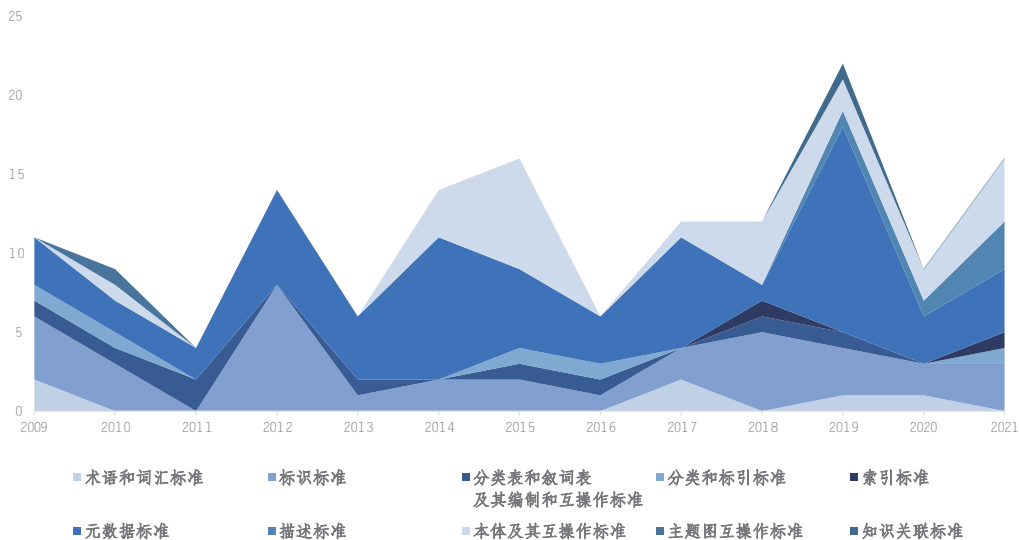


图2 知识组织领域各类别标准发布时间分布

从各类别标准的数量发展趋势来看,术语和词汇标准在2010—2016年停滞发布,其他年份的平均发布量不足两个,可以看出术语和词汇标准发展态势较为低迷;标识标准除2011年外,每年都会发布新标准,其中2012年发布了8个标准,达到峰值,2013年至今所发布的标准数量分布较为均匀,与其他类别标准相比,标识标准发展较为活跃;分类表和叙词表及其编制和互操作标准发展趋势平缓,平均每年发布1—2个新标准,但最近连续两年未发布或更新相关标准;分类和标引标准发展存在断续情况,2009年与2010年各发布1个标准后,接下来的四年时间都未有发展,2015年及以后的情况与此相似;元数据标准发展趋势较为起伏,2014年与2019年是元数据标准发布的峰值年,比较而言,元数据标准是各类别标准中数量最多、发展态

势最活跃的一类标准,年平均发布标准数接近5个;描述标准在2019年才开始发布,在此之前国内外对描述标准涉及较少;本体及其互操作标准在2009—2015年逐步发展,2015年发布或更新标准7个,达到峰值,2015年之后发展平缓,年平均发布量约为两个。

2.2.4 标准状态与标龄分析

标准的标龄是指自标准实施之日起,至标准复审重新确认、修订或废止的时间。标龄与标准状态相关,标准状态有被代替、废除和现行3种。被代替是原有的标准已失效,对该标准进行了重新确认、审批和修改,产生了新的标准;废除是指标准已不再符合要求,被彻底废止;现行是指标准当前正处于实施阶段。表2是本次调研标准的状态分布情况,现行标准共有151个,废止17个,被代替13个。术语和

标识标准、分类和标引及索引标准、元数据及描述标准在调查期间出现废止与被代替标准,而部分类别标准存在废止或被代替标准缺失

的情况,如类表和词表及其编制和互操作标准没有废止标准,关系网络构建标准仅查找到现行标准 26 个。

表 2 知识组织领域标准状态分布

标准类别	现行标准	废止标准	被代替标准
术语和标识标准	42	12	5
类表和词表及其编制和互操作标准	10	0	2
分类和标引及索引标准	6	2	5
元数据及描述标准	67	3	1
关系网络构建标准	26	0	0
总计	151	17	13

在 ISO 的标准化导则^[16]以及我国的《国家标准管理办法》^[17]中均明确规定标准实施 5 年要进行复审,即标准的“标龄”一般应为 5 年。现行标准标龄分布如图 3 所示,其中标

龄在 5 年内的标准数量最多(59 个),5 年至 10 年标龄的标准有 54 个,10 年以上标龄的标准仍有 38 个,说明目前现行标准存在“超龄”现象。

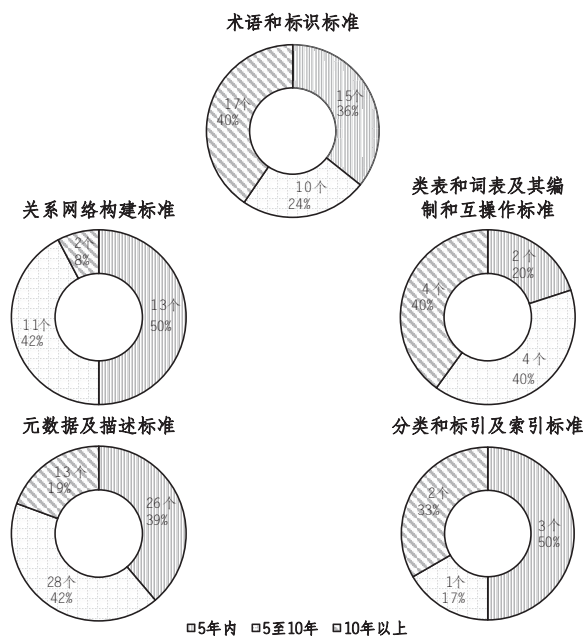


图 3 知识组织领域现行标准标龄分布

从图 3 可以看出,不同类别标准标龄分布情况存在较大差异。术语和标识标准总体而言标龄较长,5 年及以上标准数量达到 64%,标准复审时,已及时对部分标识符相关的标准进行

了审批修改,比如当下网络资源和新媒体快速发展,信息资源内容和形式越来越多样化,与之相关的旧版标准,如 GB/T 3469—2013《信息资源的内容形式和媒体类型标识》和 ANSI

MH10. 8. 2—2006《数据标识符和应用标识符标准》相继被代替;部分涉及数据库设计术语和词汇的标准,如 ANSI/INCITS/ISO/IEC TR 9007:2010《信息处理系统—概念模式和信息库的概念和术语》和 BS ISO 30300:2011《信息与文献—记录管理系统—基础和词汇》等被废止;类表和词表及其编制和互操作类别中有 80% 的标准标龄超过 5 年,现行标准大多是首次发布、经审查后仍实施至今的标准,仅部分年限已久的标准,如 GB/T 15417—1994《文献 多语种叙词表编制规则》和 GB/T 13190—1991《汉语叙词表编制规则》被修正更新;分类和标引及索引标准中半数近 5 年发布;元数据及描述标准的标龄大多在 10 年以内,废止与被代替标准数量较少,部分针对元数据管理的标准,如 ISO 23081-2:2009《信息和文献—记录元数据管理—第 2 部分:概念性和实施细则》和 BS ISO 23081-2:2009《信息与文献—记录用管理元数据—概念和实施问题》被废止;关系网络构建类别中有 50% 的标准标龄在 5 年之内,整体标龄相对于其他类别标准较短,这些标准所对应的本体、主题图与知识关联等主题在近几年发展迅速,对相应的标准规范的需求也随之提升,因此近 5 年

相关标准发布较为集中。

总体而言,知识组织领域标准“老龄化”情况较为严重,10 年以上的标准超过 1/4。聚焦到我国而言,2019 年,国家标准制修订流程发生重大调整,通报了国家标准制定流程改革的有关情况,同时修改发布了一批重要国家标准,因此我国标准标龄多聚集在 5 年内(5 年内标龄的标准 26 个;5—10 年标龄的标准 7 个;10 年以上标龄的标准 8 个),标准复审与更新都较为及时。

2.3 知识组织领域标准的引用关系分析

2.3.1 标准整体引用关系分析

在引用关系分析部分,以 2.2.2 节提到的 ISO 和 GB 标准为基础,爬取这些标准的引用和被引数据,数据均从国家标准馆获得,经去重后,共有 680 个节点和 903 组引用关系,利用 Cytoscape 软件进行可视化,结果如图 4 所示,箭头指向标准为被引标准。ISO 和 GB 标准整体引用关系比较复杂,其中 ISO 标准引用关系比较集中,图 4 左侧双圆环部分大都是 ISO 标准的整体引用关系,右侧是局部放大的引用关系(下文的图 5、6、8 也是如此)。

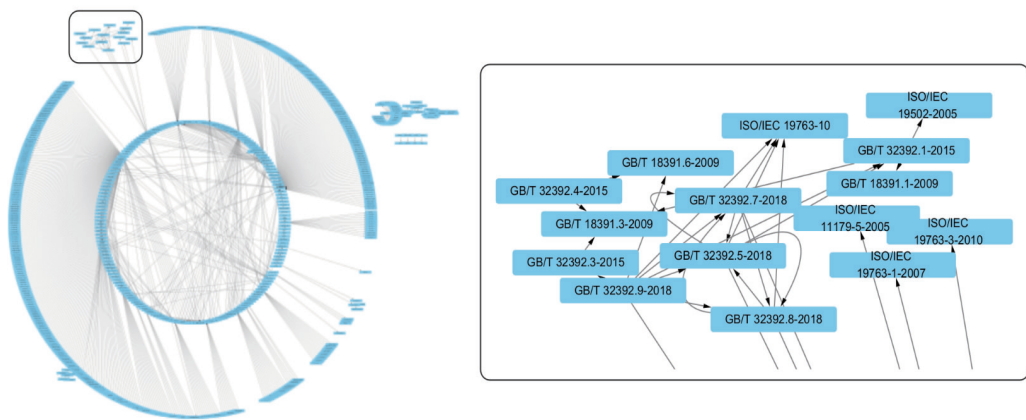


图 4 知识组织领域标准整体引用关系

2.3.2 ISO 标准局部引用关系分析

对 ISO 标准引用关系中的特殊点进行提取,分别得到 3 个子网络:引用最多的 ISO 标准、被

引最多的 ISO 标准和共被引最多的 ISO 标准。

(1) 引用最多的 ISO 标准

知识组织领域 ISO 标准中,引用最多的标

准是 ISO 5127:2017《信息与文献—基础和词汇》,共引用了 161 个标准,其引用关系如图 5 所示。该标准不仅引用了 ISO 标准中的 ISO/IEC 2382-4:1999《信息技术—词汇—第 4 部分:数据的组织》等标准,还引用了国际电工委员会(IEC)、电气与电子工程师协会(IEEE)等

多个标准化机构或组织发布的标准,包括针对文化遗产保护的术语、信息技术中数据组织的词汇、图书情报工作中国际图书馆统计学相关知识以及电影术语等多个领域的标准,跨领域引用丰富。

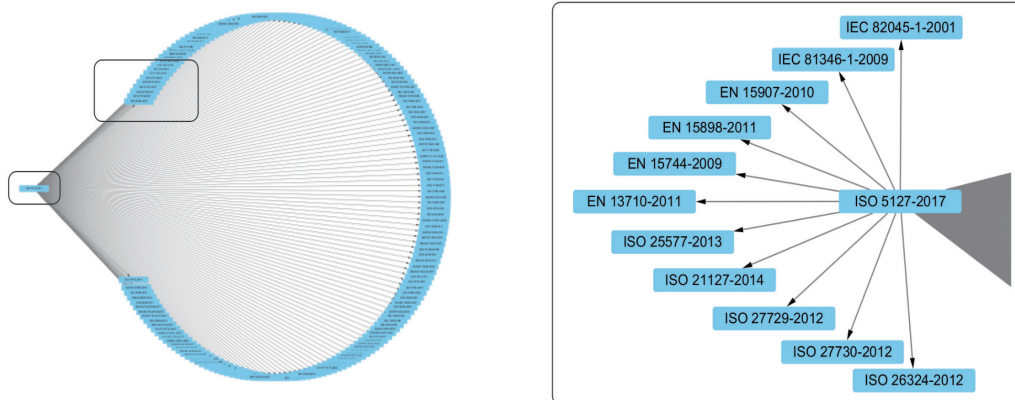


图 5 引用最多的 ISO 标准引用关系

(2) 被引最多的 ISO 标准

知识组织领域 ISO 标准中,被引最多的标准是 ISO 19115-1:2014《地理信息—元数据—第 1 部分:基础》,其引用关系如图 6 所示,该标准一共被 109 个标准引用,具有很强的通用性。在引

用该标准的标准中,以地理信息相关标准为主,如 ISO 19150-2:2015《地理信息—本体—第 2 部分:网络本体语言(OWL)的本体发展规则》等,此外还包括船舶与海洋技术、寻址、使用坐标系的标准表示法等领域的标准。

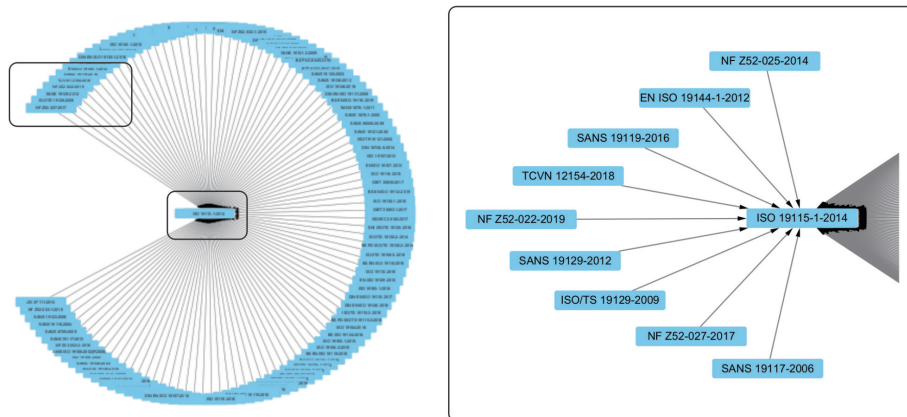


图 6 被引最多的 ISO 标准引用关系

(3) 共被引最多的 ISO 标准

知识组织领域 ISO 标准中,共被引最多的标准是 ISO/IEC 11179-3:2013《信息技术—元数据的注册(MDR)—第3部分:注册元数据模式和基本属性》、ISO/IEC 11179-1:2015《信息技术—元数据的注册(MDR)—第1部分:框架》、ISO/IEC 11179-5:2015《信息技术—元数据的注册(MDR)—第5部分:命名原则》,此3项标准都是针对元数据注册(MDR)的一些基础性标

准。如图7所示,这3项标准共被18个标准引用。引用此3项标准的标准不仅有ISO标准,还有一些国家或地区标准,如美国的 ANSI ISO 19115-2003(R2008)《地理信息—元数据》、英国的 BS PD IEC/TS 62720:2017《计算机处理测量单位标识符》、澳大利亚的 AS ISO/IEC 11179.1-2005《信息技术—元数据登记(MDR)—框架》等标准。说明这3项标准具有很高的国际通用性。

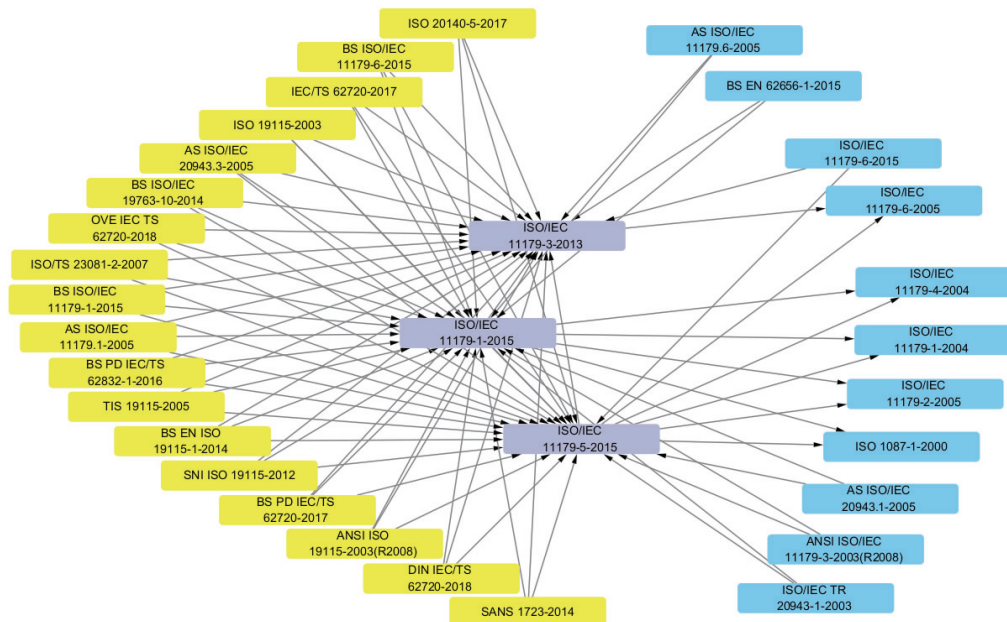


图7 共被引最多的 ISO 标准引用关系

2.3.3 GB 标准引用关系分析

(1) GB 标准整体引用关系分析

提取出 GB 标准的引用和被引关系,共有 164 个节点和 177 组关系,如图 8 所示。可发现知识组织领域 GB 标准的引用关系呈明显的分块聚集状。整体引用关系可分成 7 个分块:①号分块以 GB/T 37058—2018《图书馆编码标识应用测试》为中心;②号分块以 GB/T 32392 类标准《互操作性元模型框架(MFI)》为主,包括 GB/T 32392.1—2015、GB/T 32392.4—2015、GB/T 32392.5—2018 等;③号分块以

GB/T 3860—2009《文献主题标引规则》、GB/T 15418—2009《档案分类标引规则》的被引关系为主;④号分块是多组复杂引用关系的综合体;⑤号分块以 GB/T 4894—2009《信息与文献术语》为中心,其中有 6 个标准引用了该标准;⑥号分块以 GB/T 2901—2012《信息与文献 信息交换格式》为中心,包括基于该标准的 3 组引用和 3 组被引关系;⑦号分块则是多组引用关系较少的标准,包括 7 组一对一引用关系,1 组一对二引用关系和 1 组一对三引用和被引关系。

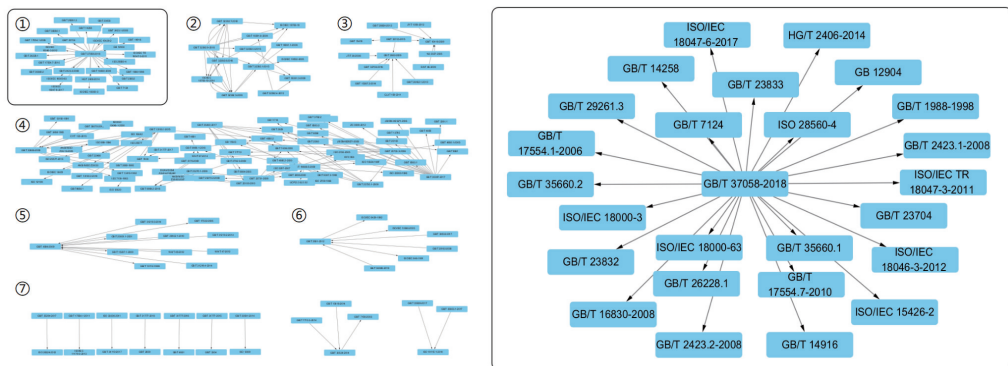


图8 GB标准整体引用关系

(2) GB标准局部引用关系分析

本文主要对①号分块和④号分块进行分析,①号分块 GB/T 37058—2018 是 GB 标准中引用最多的标准,④号分块则是引用关系相对复杂的分块。

其中,①号分块以 GB/T 37058—2018《图书馆编码标识应用测试》为中心,共有关于该

标准的 25 组引用关系,如图 9 所示。该标准主要对图书馆图书编码标识、图书馆层架编码标识、图书馆读者证编码标识以及图书馆射频识别(RFID)技术和标签做了规范性要求。此标准不仅涉及标识和编码类知识,还涉及多项技术知识,因此其引用标准较多,跨领域引用也较多。

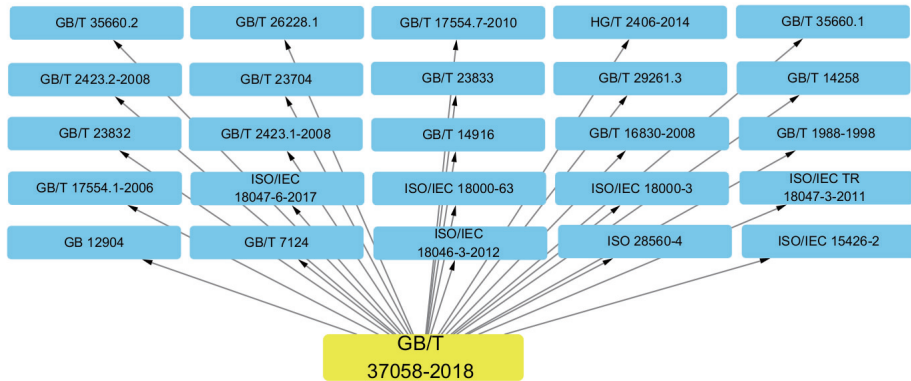


图9 引用最多的GB标准引用关系

对④号分块关系数大于等于 2 的节点进行提取,提取后的子网络共有 25 个节点和 36 组关系,如图 10 所示。在这个子网络中被引最多的标准是 GB/T 7408—2005《数据元和交换格式信息交换 日期和时间表示法》、GB/T 4880.2—2000《语种名称代码 第 2 部分:3 字母代码》,这两个标准均是比较基础和通用的标准。在国家标准馆网站对两项标准进行检索,分别有 394 个

和 100 个被引记录。其中,GB/T 7408—2005 被知识组织领域的 GB/T 35397—2017《科技人才元数据元素集》等 4 项标准引用。此外,中国标准视听作品号相关标准 GB/T 23730.1—2009 与 GB/T 23730.2—2009 共同引用了 GB/T 11714、GB/T 17710 等 4 项标准,此两项标准相似度相对较高。

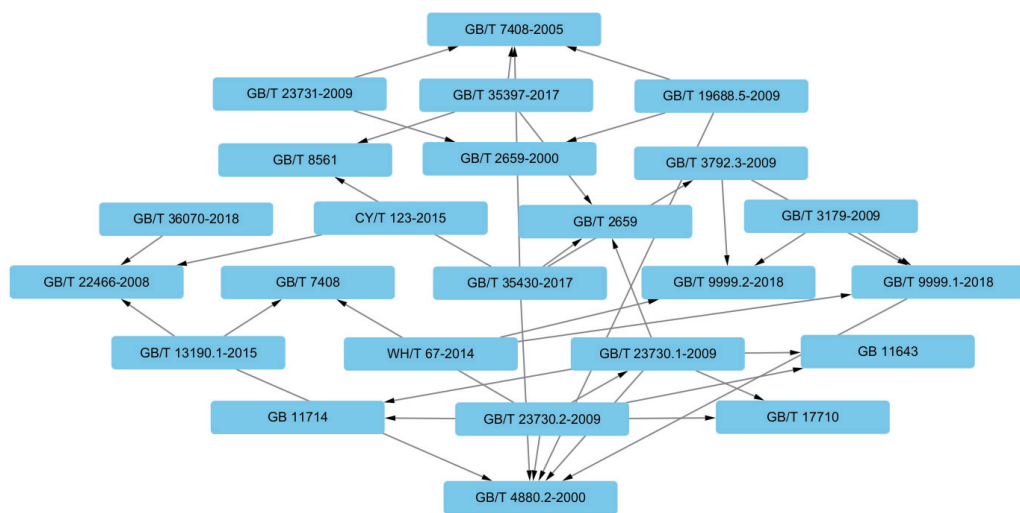


图 10 引用关系复杂的 GB 标准引用关系

3 知识组织领域标准规范的重要变化与特点

3.1 知识组织领域标准规范的重要变化

新的信息环境下,作为知识组织对象的资源呈现出传统资源与数字资源并存、多源异构等特点,需要统一、规范、标准化管理。为适应知识组织体系的变化,相应的标准规范也在不断地发布、修订和更新,其重要变化主要体现在以下几个方面。

3.1.1 资源描述标准更新

知识组织领域标准规范的重要变化之一是资源描述标准的更新,主要体现为以下两方面。

(1) 新版 RDA 上线。《资源描述与检索》(Resource Description and Access, RDA)是以 FRBR 和 FRAD 概念模型为基础框架的编目规则,于 2010 年以工具套件(RDA Toolkit)的形式正式发布,目的是为了满足数字环境下的资源著录与检索需求。2019 年 10 月, RDA 指导委员会发布了“3R 项目完成声明”,经过全新修订的 RDA 在 2020 年 12 月 15 日正式上线。新版 RDA 在内容和体例编排上均有较大变动,其中,

内容方面的变化主要体现在编制范围、实体、属性和关系等方面。比如,新版 RDA 不包含具体的主题编目规则,但提供了一组广泛的关系元素,用来对作品和作为其主题的 RDA 实体之间进行关联;除此之外,新版 RDA 共定义了 13 个实体,即 RDA 实体(一个超类实体,用于反映 RDA 中实体的共同特性)、作品、内容表达、载体表现、单件、行为者、个人、集体行为者、团体、家族、命名、地点、时间段,与 13 个实体相联系的属性和关系均被称为“元素”,将关系说明语当作关系元素,连同属性元素对实体进行描述^[18]。

(2) 中国国家标准 GB/T 3792—2021《信息与文献 资源描述》的发布。国际标准书目著录 ISBD 系列和英美编目条例 AACR2 服务于图书馆文献编目已长达半个多世纪,我国文献著录国家标准最初发布于 1983 年。当前网络环境下信息资源数量激增、类型多样、更新速度加快,使得传统的著录规则已不能满足语义网环境下资源描述和组织的需要。为应对社会环境的变化, IFLA 制定了一些标准规范,比如 1998 年发布的 FRBR,详细梳理了各实体的属性及实体之间的关系;2010 年, RDA 正式发布;2011 年, ISBD 统一版发布。此后,我国信息与文献

标准化技术委员会开始将更新现有文献著录国家标准提上重要议程。2012 和 2013 年,全国信息与文献标准化技术委员会两次召开研讨会,深入讨论我国文献著录国家标准研制的关键性和方向性问题。2021 年 3 月,中国国家标准 GB/T 3792—2021《信息与文献 资源描述》发布,并于 2021 年 10 月 1 日起实施。该标准依据 ISBD 统一版,将各资源类型著录规则合并成一个统一的标准,代替了 GB/T 3792.1—2009《文献著录 第 1 部分:总则》等 GB/T 3792 系列标准,是一个面向各种类型资源的通用资源描述标准。

3.1.2 标识标准被全面修订

标识在出版、图书情报等领域常被用来作为识别、区分以及定位信息与文献的专用符号,比如国际标准书号、国际标准刊号,它们可以作为支持版权管理的基础。近年来,数字资源成为新的信息文献产品形态,不同属性的资源之间相互关联、复合呈现,传统的面向实物对象的标识符已不能满足实际需求,需要扩展到对网络对象进行标识。同时,网络资源之间错综复杂的关系对标识符提出多元化的需求,比如 2015 年 ISO 发布的 ISLI,不再标识一个具体的物,而是标识实体之间的关联关系。信息资源标识的原则一直是 ISO 所属的 TC46(信息与文献标准化技术委员会)和 SC9(识别与描述分技术委员会)关注的重点。2018 年,ISO 针对现代媒体融合环境下“标识原则”问题展开专项研究,研究成果以技术报告《标识原则》的形式发布。该文件对 ISO 标识符制修订工作提出了明确的要求。目前,国际标准 ISBN 和 ISRC 都完成了新一轮修订,ISBN 第五次修订(2017 年发布)的主要内容是在数字出版物中加入数字版权管理(DRM)后 ISBN 的分配原则以及数字环境下 ISBN 解析的方法。ISRC 修订(2019 年发布)的内容涵盖了分配编码的结构,以及分配、管理、编码检索服务等,阐述了 ISRC 码的格式,确保代码唯一性的机制以及为录音分配 ISRC 码的实现方式。中国国家标准 GB/T

3469—2013《信息资源的内容形式和媒体类型标识》被 GB/T 3792—2021《信息与文献 资源描述》替代。

3.1.3 大数据标准化工作取得进展

当前信息环境下,与知识组织相关的大数据标准化工作受到关注。自 2012 年起,ISO/IEC 陆续开展大数据标准化工作,目前已发布了多项大数据标准。比如,2019 年发布了 ISO/IEC 20546:2019《信息技术—大数据—概述和词汇(第一版)》,2020 年发布 ISO/IEC 20547-3:2020《信息技术—大数据参考体系结构—第 3 部分:参考体系结构(第一版)》。近年来,中国在大数据标准化建设方面也取得了较大进展。2014 年,全国信息技术标准化技术委员会大数据标准工作组在北京成立,其中心工作是推进大数据、元数据、数据分类、数据开放共享等方面的标准化,以促进相关领域产业、应用和服务等各方面的稳步发展。2020 年,中国国家标准 GB/T 38667—2020《信息技术 大数据 数据分类指南》、GB/T 38664.1—2020《信息技术 大数据 政务数据开放共享 第 1 部分:总则》、GB/T 38664.2—2020《信息技术 大数据 政务数据开放共享 第 2 部分:基本要求》、GB/T 38664.3—2020《信息技术 大数据 政务数据开放共享 第 3 部分:开放程度评价》等系列标准发布,其归口单位是全国信息技术标准化技术委员会。除了国家层面以外,地方也在大力推进大数据的标准化工作,2017 年国内首个大数据标准化技术委员会在贵阳成立,主要负责贵州大数据领域标准体系建设及地方标准制定、修订、技术审查工作。中国的大数据标准化也在有序推进,不断取得新进展。

3.2 知识组织领域标准规范的特点

3.2.1 标准规范体系日益完善

知识组织领域现行的标准规范中既有常见的叙词表编制标准、术语与词汇标准、元数据标准、标识与索引标准,也逐渐出现了新的知识关联、主题图互操作等关系网络构建标准规范,标

准类型呈现多样化趋势,标准规范体系日益完善。

3.2.2 元数据标准成为知识组织标准规范的重心

元数据标准是描述某类资源的具体对象时所有规则的集合。元数据标准的发展相对成熟,近13年知识组织领域的元数据标准数量最多,并在不断修订中。多源异构信息资源的快速增长对元数据标准提出了更多要求。2020年ISO发布的描述型元数据国际标准将其范围扩大到包括都柏林核心元数据倡议(DCMI)元数据术语中的所有属性和分类。除了描述型元数据标准以外,关于元数据交换、元数据模块的技术互操作性的系列标准规范也不断发布和修订。近年来,科学数据共享元数据内容标准、开放数据元数据标准、大数据环境下的文献元数据标准规范建设开始受到重视。

3.2.3 本体及其互操作标准呈现强劲的发展势头

本体是用于描述或表达某一领域知识的一组概念或术语,具有良好的概念层次结构,能实现对逻辑推理的支持。随着本体构建与描述实践的日益丰富,本体的相关标准在近几年迅速涌现,从特定领域的本体到本体互操作以及本体描述语言,呈现良好的发展势头。这些标准能体现新技术的应用,包括在Web本体语言中开发本体的规则、互操作性元模型框架中的本体注册元模型等。最新标准是ISO于2021年发布的ISO/IEC 21838-1:2021《信息技术—顶级本体—第1部分:要求》,规定了与领域无关的顶级本体(TLO)所需的特征,该本体可与较低级别的领域本体一起使用,以支持数据交换、检索、发现、集成和分析,它还支持与实现语义互操作性相关的各种其他目标。

3.2.4 新技术应用类相关标准相继出现

大数据、区块链、人工智能技术发展迅猛,知识组织领域关于新技术应用的标准规范也相继出现。目前已有一些现行标准,同时还在起草其他标准。现行标准中有ISO发布的关于大

数据的参考架构、概述和词汇、用例和派生需求的标准,也有关于领域知识图谱构建技术规程、技术要求的团体标准。2021年ISO还发布了ISO/TS 23258:2021《区块链和分布式账本技术—分类和本体》,指定了区块链和分布式账本技术(DLT)的分类法和本体。这在一定程度上预示着知识组织领域标准规范的未来发展趋势。

3.2.5 中国标准对国际标准的同步性和引领性有待加强

本研究统计的41个中国现行国家标准中有20个是采用的ISO标准。中国标准的修订具有一定的滞后性,与ISO标准对应的同名或同类型的中国标准大都在4—5年后才发布,GB/T 3792—2021《信息与文献 资源描述》更是晚于国际标准近10年。此外,中国一些现行标准采用的是已被废止的ISO标准,如与GB/T 25100—2010《信息与文献 都柏林核心元数据元素集》相关的ISO标准在2017年和2019年分别对第一部分和第二部分进行了新的修订,已废止2010年的版本。由此可见,中国国家标准与国际标准的同步性有待加强。另外,我国在其他领域制定的标准已有引领国际标准的趋势,比如2018年由中国信息通信研究院提出的ITU-T F.743.20《大数据基础设施评测框架》和ITU-T F.743.21《数据资产管理框架》两项国际标准正式立项。相比之下,在知识组织领域,中国对国际标准的贡献还不突出,需要不断加强。

4 未来发展趋势

通过对近13年国内外知识组织领域的标准规范及其发布、更新及修订情况进行分析,可以预测未来有以下几个发展趋势。

(1)知识组织领域标准规范呈现交叉融合趋势。在数字化环境下,知识组织各个细分领域的界限逐渐被打破,不同类型的标准规范相互参考、相互渗透,知识组织的标准植根于网络技术标准中,比如资源描述标准中融入语义网

标准体系,信息技术标准中融入分类和本体的相关标准,等等。此外,多个标准整合成一个统一标准,形成通用标准也是一种可见的发展趋势。另一方面,知识组织领域标准规范的交叉融合还体现在跨行业借鉴和应用中,如 ISLI 最初在新闻出版行业被提出并广泛应用,大数据分类标准规范在科技、工业、政务等各个行业发挥重要作用。

(2) 知识组织领域标准规范的建设机制将更加灵活。从目前发展情况来看,知识组织领域标准的制定和推广并非仅靠专门的国际标准组织和国家标准组织,能够快速适应技术发展进程的行业标准、地方标准、团体标准正在不断出现。如 W3C 制定的一系列标准, RDA 完全是靠技术先进性引领。未来,标准规范的更新速度更快、制定周期缩短,建设机制更加灵活。

(3) 知识关联、数据交换将是未来知识组织标准规范的关注重点。面对海量异构的信息资源,人们越来越需要实现智能化检索。要满足这一需求,就必须实现不同类型知识组织体系间的互操作,这种互操作既包括数据交换问题,也包括知识关联问题。近年来,国际标准、中国

国家标准、美国国家标准中相继发布了用于互操作和保存的数据交换协议、XML 元数据交换(XMI)、知识关联通用规则等相关标准,这些将成为未来标准内容关注的重点。

5 结语

本文通过对近 13 年来国内外知识组织领域标准规范数据的收集和分析,系统梳理了知识组织标准规范的发展现状,总结了当前阶段呈现的变化与特点。可以看出,我国制定和发布的信息组织标准规范在数量上与国际组织以及其他发达国家逐渐持平,说明我国知识组织领域标准化工作取得了显著进展。同时,知识组织领域的标准规范逐渐呈现出交叉融合、建设机制更加灵活、知识关联与数据交换逐渐成为重点的发展趋势,这也为未来我国相关标准规范的制定提供了启示:既要根据知识组织领域的发展需求对传统标准规范进行不断更新,也要探索新技术在知识组织领域的应用标准;既要与国际标准保持同步,又要力争发挥引领作用。

参考文献

- [1] ZENG M, MARGIE H, JIAN Q, et al. Knowledge organization systems(KOS) standards[J]. Proceedings of the American Society for Information Science and Technology, 2008, 44(1): 1-3.
- [2] 刘华. 国内外知识组织体系标准的现状及发展趋势研究[J]. 情报杂志, 2011, 30(3): 14-18. (LIU H. Status quo and development trend of knowledge organization system standards at home and abroad[J]. Journal of Intelligence, 2011, 30(3): 14-18.)
- [3] 卜书庆. 近年来国际知识组织规范的进展研究[J]. 中国图书馆学报, 2010, 36(5): 69-74. (PU S Q. On the progress of international knowledge organization standards in recent years[J]. Journal of Library Science in China, 2010, 36(5): 69-74.)
- [4] 杨雨霖. 我国图情档领域信息组织标准发展现状研究[J]. 图书馆学研究, 2014(11): 63-66. (YANG Y L. Research on the development status of information organization standards in the field of library and information archives science in China[J]. Research on Library Science, 2014(11): 63-66.)
- [5] 张琳. 信息组织领域国内外标准规范内容比较研究[J]. 情报杂志, 2016, 35(3): 185-189. (ZHANG L. A comparative analysis of standards contents in the area of information organization[J]. Journal of Intelligence, 2016, 35(3): 185-189.)
- [6] 张琳. 信息组织领域标准规范开发管理现状比较研究——以 ISO、美国、英国和中国为例[J]. 图书馆杂

- 志,2017,36(9):68-74. (ZHANG L. Comparative analysis of standards development and management in the field of information organization;take ISO,NISO,BSI and SAC as examples[J]. Library Journal,2017,36(9):68-74.)
- [7] NEEDLEMAN M. The resource description and access standard[J]. Serials Review,2008,34(3):233-234.
- [8] ESCOLANO R E. Consolidated edition of ISBD,International Standard Bibliographic Description;a standard to trust,a quality brand[J]. Cataloging & Classification Quarterly,2014,52(8):835-854.
- [9] 宋文,朱学军.《资源描述》国家标准及对我国信息资源描述标准体系的思考[J]. 数字图书馆论坛,2016(12):21-27. (SONG W,ZHU X J. Resource Description national standard and thoughts on information resource description standard system[J]. Digital Library Forum,2016(12):21-27.)
- [10] 杨波,胡立标. 用于社会科学信息组织的元数据标准——DDI[J]. 数据分析与知识发现,2005(8):7-11. (YANG B,HU L Y. Metadata standard for social science information organization—DDI[J]. Data Analysis and Knowledge Discovery,2005(8):7-11.)
- [11] KALITA D,DEKA D. Searching the great metadata timeline;a review of library metadata standards from linear cataloguing rules to ontology inspired metadata standards[J]. Library Hi Tech,2020,39(1):190-204.
- [12] 贾大林. ISLI 在知识服务中的价值和应用研究[J]. 出版参考,2019(7):5-7. (JIA D L. Research on the value and application of ISLI in knowledge service[J]. Publishing Reference,2019(7):5-7.)
- [13] 朱江,李欣怡,任晓亚,等. 国际标准关联标识符在图书馆的应用研究[J]. 图书馆建设,2020(2):56-61,69. (ZHU J,LI X Y,REN X Y,et al. International Standard Link Identifier and its application in library[J]. Library Development,2020(2):56-61,69.)
- [14] 曹树金,王志红,王连喜. 国内外知识组织研究内容与发展——基于《图书情报工作》与 Knowledge Organization 期刊论文的比较分析[J]. 图书情报知识,2017(4):100-112. (CAO S J,WANG Z H,WANG L X. The research and development of knowledge organization in China and abroad;a comparative review study of Library and Information Service and Knowledge Organization[J]. Document,Information & Knowledge,2017(4):100-112.)
- [15] HIDER P. The terminological and disciplinary origins of information and knowledge organization[J]. Education for Information,2018,34(2):135-161.
- [16] ISO/IEC directives,part 2:rules for the structure and drafting of international standards[S]. ISO/IEC,2004.
- [17] 国家市场监督管理总局. 国家标准管理办法[EB/OL]. (2022-09-09)[2022-11-08]. <https://baijiahao.baidu.com/s?id=1744689801039690045&wfr=spider&for=pc>. (State Administration for Market Regulation. National standards management measures[EB/OL]. (2022-09-09)[2022-11-08]. <https://baijiahao.baidu.com/s?id=1744689801039690045&wfr=spider&for=pc>.)
- [18] 丁政,蔡丹. 新版 RDA 探微——3R 项目稳定阶段略述[J]. 数字图书馆论坛,2020(10):32-37. (DING Z,CAI D. A probe into the new version of RDA;a brief introduction to the stable phase of the 3R project[J]. Digital Library Forum,2020(10):32-37.)

曹树金 中山大学信息管理学院教授,博士生导师。广东 广州 510006。

曾盈盈 中山大学信息管理学院硕士研究生。广东 广州 510006。

廖赛源 中山大学信息管理学院硕士研究生。广东 广州 510006。

曹茹桦 中山大学信息管理学院博士研究生。广东 广州 510006。

(收稿日期:2022-03-17;修回日期:2022-11-08)