

DCMI 2021虚拟年会综述*

范炜 林君松

(四川大学公共管理学院, 成都 610064)

摘要: DCMI年会是元数据领域重要的国际学术研讨盛会。根据DCMI 2021虚拟年会资料, 归纳总结出当前国际元数据学界与业界的主题聚焦、项目经验分享及技术应用等。此次年会关注语义网、关联数据和本体的技术赋能, 分享了芬兰国家本体与语义门户建设、Schema.org与DC元数据的融合发展等创新思路与经验做法。在元数据应用领域方面, 主要涉及文化遗产元数据、书目资源转换升级、隐私保护、科学数据管理等。在元数据技术系统方面, 讨论了人工智能与自动标引、实体识别与管理、关联数据发布与资源管理平台等。跟踪DCMI年会最新进展, 可为中国本土元数据研究与应用实践提供参考借鉴。

关键词: DCMI; 2021虚拟年会; 元数据创新; 语义网; 关联数据; 本体; 多元化

中图分类号: G254 **DOI:** 10.3772/j.issn.1673-2286.2022.05.011

引文格式: 范炜, 林君松. DCMI 2021虚拟年会综述[J]. 数字图书馆论坛, 2022 (5) : 66-72.

DC元数据 (Dublin Core Metadata) 是全球范围内影响力最大、应用范围最广的网络资源通用描述标准。都柏林核心元数据组织倡议 (Dublin Core Metadata Initiative, DCMI) 制定了DC元数据标准, 并持续推动元数据的应用实践。

1995年, DCMI在美国俄亥俄州都柏林发起第一次元数据研讨会。自此, DC元数据吸引了全球范围内众多信息专业人员参与其中, 产生了广泛的影响力。DCMI逐渐从区域性会议转变为国际性年会, 鼓励全球范围内对元数据感兴趣的人员参与进来分享实践经验与看法。自2001年起, DCMI不再使用DC元数据研讨会 (Dublin Core Metadata Workshop) 提法, 不限制参与人数, 开始以年份命名会议, 目前已连续举办29次, 已发展成为全球最具影响力的元数据学术盛会。

由于新型冠状病毒肺炎疫情肆虐全球并产生持续影响, 2020年起DCMI年会由线下实体会议转变为线上网络会议。疫情虽然阻隔了面对面的会议交流, 但同时推动了在线会议交流形式的灵活性与多样性, 吸引全球范围更广泛的关注与参与。

DCMI 2021虚拟年会于2021年10月4—15日成功

举办, 历时12天, 其规模和体量较以往更大, 是至今为止元数据年会中规模最大的一次。会议汇报交流者来自全球26个国家的100多名专家学者、专业人员以及高校师生等。DCMI 2021虚拟年会主题为“元数据创新” (Metadata Innovation), 探讨全球视野的元数据创新发展, 分享元数据实践成果与项目经验^[1]。此次年会组织者阵容强大, 会议主席是美国肯特州立大学的曾蕾教授, 会议组委会成员包括我国武汉大学的黄如花教授、韩国成均大学的Sam Oh教授等。

此次年会形式多样, 包括主旨报告、主题汇报、最佳实践、特邀讲座、培训讲座和会员论坛等多种形式。一个创新之举在于, 增设了学生论坛。通过不同国家地区学生的学术成果分享, 提供与业届专家交流机会, 为元数据专业人才培养、新生力量的培养提供了交流与展示舞台。

根据年会日程, 通过DCMI网站采集参会者汇报内容的摘要与标题, 整理形成文本文件。借助Python的分词包jieba, 通过段落、语句切分, 自定义停用词表和用户词典, 经过数据清洗与分析, 统计词频在3个以上的关键词共有310个。除元数据外, 数字化、数据和信

* 本研究得到四川大学创新火花重点项目“巴蜀非物质文化遗产知识图谱构建研究” (编号: 2018hhf-18) 资助。

息等通用词词频较高,其次是图书馆、档案馆、网络、项目、政策、科研、文化遗产、数据建模、关联数据及本体等。

此次年会内容丰富且交流形式多样,限于篇幅,本文首先重点分析特邀主旨报告,随后按研究主题选择性进行评述,以期为国内读者提供一份元数据研究线索资料。

1 年会特邀主旨报告分析

DCMI 2021虚拟年会特邀两位资深专家做主旨报告。一位是学者型专家,芬兰赫尔辛基大学的Eero Hyvönen教授;另一位是实践型专家,来自谷歌的Dan Brickley目前负责Schema.org的运维。以下重点分析两份主旨报告。

1.1 芬兰国家级本体和语义门户建设经验

Eero Hyvönen教授的主旨报告主要介绍了Sampo模型及其系列语义门户项目经验,芬兰国家本体建设与服务,以及面向用户服务和数据开放重用的7星模型。

(1) 基于Sampo模型的系列语义门户构建。Sampo模型是一种用于创建跨领域本体和关联数据集成的元模型,是语义门户建设的核心所在^[2]。Sampo模型主要包括三部分:用于创建和发布关联数据的业务模型;面向最终用户的资源内容视图;基于视图的分面检索和数据分析。其中,业务模型依赖于国家本体基础设施提供的本体服务,协调多个来源的内容提供,采用国际标准与常见做法,以语义关联的方式在应用中集成各类资源。在面向最终用户使用的语义平台服务,基于Sampo模型的语义门户从提升用户体验出发,除提供传统资源的文本检索、浏览与下载等服务外,还提供了多样化的内容视图,帮助用户从多维度进行语义检索与分析。在Sampo模型基础上,开发了多个专题语义门户,如WarSampo、CultureSampo、HistorySampo、LawSampo等^[3]。

(2) 国家本体模型基础设施建设。芬兰国家级本体基础设施建设是较为特殊的一种集中式本体项目实施做法。Eero Hyvönen以芬兰博物馆项目建设所面临的问题和挑战,如多源异构数据发布问题、本体无法适应语义网应用程序、海量数据索引效率,以及开发大型跨领域词表所需要的领域专家的参与问题等,提出面向国家级本体基础设施项目FinnONTO(2003—2012)。

该国家本体基于叙词表构建共享本体,能够实现不同领域本体之间的自动关联与语义丰富。由特定领域专家组通过分布式参与方式进行领域本体开发。通过协调芬兰国内众多机构和专家学者,开发一系列领域本体,主要包括通用概念、参与者、地点、时间、事件以及特定主题领域,如医疗健康和农林等。在分布式参与模式下,不同领域专家通过词表对齐与转换生成本体。最后,统一映射到YSO上层本体,实现不同领域本体之间的关联映射和语义互操作。各个本体汇聚成一个整体,即KOKO本体,通过ONKI(onki.fi)本体服务平台,提供本体发布、编辑与维护等^[4]。

(3) 数据开放重用的7星模型。数据的最终价值是服务于用户需求,通过借助网络能在最大程度上被更广泛范围的用户使用并重用。以用户为中心提供数据服务并兼顾数据发布者的应用需求,Eero Hyvönen教授在关联数据发布5星模型基础上,创新性扩充了2星要求,形成关联数据服务7星模型。6星要求为数据提供模式和文档,以便用户能够理解和重用数据。7星要求验证数据并指出其来源,以使用户可以信任数据质量^[5]。Eero Hyvönen认为,5星数据能够吸引用户使用这些数据,但并不足够支持用户“消费”这些数据。关联数据服务7星模型的提出,一方面,增进了用户对发布数据的内容结构、质量、来源等多方面的了解,尽可能地减少用户为使用数据所做的额外工作,刺激用户对数据的“消费欲望”;另一方面,也要考虑到数据发布者的工作负担。因此,在平衡数据用户与数据发布者二者的利益需求上,目前LDF.fi平台符合其提出的关联数据服务7星模型,支持用户上传发布自己的数据和使用各类服务。

通过早期的项目建设为后续的文化遗产保存利用和数字人文研究奠定了较为坚实的基础,将各类资源借助语义门户网站呈现,以知识单元形式提供给用户,服务于知识发现。经历了第一代只提供给用户检索、浏览的门户建设模式,向第二代为用户提供检索工具、帮助数字人文研究人员解决研究问题的模式转变,实现从为用户提供资源到为用户提供解决问题工具的过渡,逐步走向以知识发现、问题解决和自动推理等AI技术赋能的第三代门户建设^[6],推动数字人文研究发展。

1.2 DC元数据与Schema.org竞争共生发展

Dan Brickley见证了DCMI与DC元数据的发展历史,是典型的业界技术实践专家。他在DCMI 2012年会

上曾做主旨报告,“What is left to do?”,以老问题新回答的方式,讨论了初创的Schema.org与DC的关系,指出在语义网与关联数据中发展元数据^[7]。9年之后,他再次受邀作为主旨报告嘉宾,以“Schema.org and Dublin Core: the next chapter”为题,分享了最新网络元数据实践。

Dan Brickley参与制定了语义网相关标准,如W3C RDF Schema规范^[8],网络个人信息描述的事实常用元数据方案FOAF。目前其负责的Schema.org是由网络社群驱动的开放元数据标准,被主流搜索引擎公司与各类知识图谱应用所采用。

Dan Brickley首先从参与DC元数据制定谈起,结合实际工作,介绍了DC与Schema.org的发展历程与应用情境,主要内容归纳为以下4点。

(1) 跨区域的组织协作是运行保障。当前各类网络资源的数据化开发与利用,需要各类专业人员跨领域进行协作。W3C为网络数据标准的制定与推广提供了保障。通过网络社群参与的专业人员所拥有的不同学科背景,给标准的推广创造了有利条件。这些专业人员将这些数据标准或应用或拓展,使行业彼此之间的开放关联性增强。

(2) DC与语义网的强绑定是必然选择。从元数据发展历程看,DC元数据能够发展为最通用的元数据方案,实现机器的可读取和可理解,这与契合语义网发展,选择与W3C的RDF、OWL紧密结合是分不开的。各类组织机构将DC元数据用于网站资源结构化描述,通过关联数据方式发布,逐步形成开放关联的数据网络。

(3) Schema.org的数据变现能力强。Schema.org是在谷歌、必应、雅虎、Yandex这4家搜索引擎的共同支持下,由网络社群开发的一套网络资源描述与标注方案。最早是为了搜索引擎更好地识别网页内容,以微数据格式嵌入网页,赋予机器理解网页内容的能力。在此基础上,通过实体与关系的丰富化,逐步增强语义化描述能力,能够处理文章、食谱、事件、事实核查、数据集等多种结构化数据对象,实现检索优化和语义理解等^[8]。

(4) 两者共同面临的挑战。DC元数据作为国际标准(ISO: 15836)的通用扩展性强,Schema.org的网络实践活跃且扩展丰富。尽管两者在当前拥有显著影响力,但面临的问题依然存在,例如服务于社会发展的需求转变所存在的应用范围、描述粒度等问题,还需要DCMI与Schema.org社群加强交流,探索网络信息资源描述与发现的共融发展之路。

Dan Brickley的主旨报告引发了参会者对DC元数据与Schema.org融合发展方向的思考。高质量的结构化与语义化的数据资源是知识图谱与人工智能应用的重要支撑,Schema.org代表了网络资源描述与标注的新阵地,DC元数据与Schema.org的开放融合也是发展的必然选择。

2 年会研究主题评述

此次年会分享了来自不同领域的众多主题演讲与项目实践,通过不同机构组织之间的交流沟通,为元数据的创新应用提供全球视野与新发展思路。以下主要对5个主题进行线索式概要评述。

2.1 文化遗产元数据

此次年会有关文化遗产元数据的应用实践非常丰富,代表性报告评述如下。

日本筑波大学的Shigeo Sugimoto教授等在数字空间中,讨论了面向物质文化遗产和非物质文化遗产的关系表示的元数据模型^[9],提出文化资源开发利用中所面临的一系列研究问题与挑战。例如,为哪些文化实体创建元数据,烟花属于艺术活动还是物理实体,作为非物质文化遗产的传统舞蹈属于舞蹈技巧还是表演,人文学科中的研究数据是将个体还是整体档案作为一个数据集,以单件为中心的元数据描述是否适用于烟花表演、文化活动和新媒体艺术等问题。

日本IT顾问工程师Tetsuya Mihara结合自身兴趣,介绍了日本媒体艺术数据库(Media Art Database, MADB)的项目开发^[10]。他提出了一种基于单件与基于内容相结合的本体建模方法,既从资源内容建模,又从单件属性说明资源的个体特征,介绍其在数据收集、组织等过程中遇到的困难与挑战。MADB中主要包括四类资源:日本漫画、动画、视频游戏及其他新媒体艺术。目前MADB数据库中前三类资源居多,收集了日本近480 000本漫画杂志、9 000部动画和48 000种视频游戏,以关联数据发布到GitHub。

斯里兰卡科伦坡大学的Chiranthi Wijesundara从无形和短暂文化实体建模出发,介绍了文化遗产信息的丰富性与相关特征,分析了文化遗产数字环境模型CHDE,将建模对象区分为数字空间和物理空间两类,对两种不同空间范围的实体属性映射到CIDOC-CRM

和FRBR, 从而构建本体, 实现数字资源聚合。

Europeana目前包括44个国家近4 000个文化机构的文化数据资源网络。在元数据语义丰富化方面, EDM数据模型为不同机构之间的语义关联和描述提供数据模型基础。Europeana的Antoine Isaac在本次年会上介绍了基础实体收集过程和实体管理的一些做法。例如, 开发与实体关联的协作用户库、通过推荐引擎推荐新项目等。Antoine Isaac等分享了Europeana在建设过程中面临的多语言挑战, 对用户网站浏览、阅读网站内容、文本阅读和搜索4类行为进行多语言支持。多语言支持的困难体现在24种官方语言更新、每两周更新一次用户界面以及资金支持不足等。他们提出了使用谷歌自动翻译服务节省志愿者时间与资金, 翻译存疑之处寻找母语人员解决, 以及项目建设伙伴以邮件或网站反馈形式进行翻译更新。采用谷歌云翻译API对检索结果实时翻译更新等手段, 提高多语言兼容效率。

上海图书馆的刘炜副馆长、夏翠娟研究馆员介绍了上海图书馆历史人文数据平台的构建过程, 以及现有的三大数字人文平台(元数据搜索、全文提供系统和语义知识系统)。上海图书馆数字人文平台语义架构设计采用数据中台理念构建, 结合智慧图书馆建设, 强调知识服务的主体在于数据中台的建设, 数据中台能够最大程度地增强图书馆对知识单元的查找、关联、分析和可视化能力。未来智慧图书馆建设中DC将发挥核心基础性作用。

其他汇报包括: 泰国朱拉隆功大学的Songphan Choemprayong副教授分享的泰国莫科伦地区文化遗产语言建模研究; 泰国孔敬大学的Suwanee Hoaihongthong从文化遗产信息管理角度, 将壁画特征、壁画使用者信息行为特征、数据注册者使用元素与VRA CORE进行对比分析等; 瑞典索德脱恩大学的Karin Hansson副教授从开放研究数据库的角度解读当前面向文化遗产元数据与图片共享的开放研究, 指出早期对开放数据库中的文化遗产数据共享存在个人研究偏好、知识产权保护和相关元数据标准缺乏的问题, 提出在开放数据库中可以申请怎样的准则和元数据去描述开放获取的数据。

DCMI曾在2012年成立文化遗产元数据任务组。2022年初由Shigeo Sugimoto教授牵头成立核心文化元数据建模兴趣组, 开展图档博视角的文化遗产多样性数据建模研究。未来文化遗产领域的元数据研究与应用实践会持续活跃。

2.2 图书馆元数据业务实践

英国联合信息系统委员会的Neil Grindly分享了英国国家书目知识库NBK的建设经验和应对数据应用的M计划。由JISC负责建设的NBK数据库, 收集了来自176个机构约1.3亿条数据记录。这些数据包括各类图书馆、档案馆的编目数据, 以及开放存取(OA)资源数据和出版社数据等, 经过去重、对比等处理得到4 800万条记录, 这些数据对外开放并发布在WorldCat中, 便于编目处理与分析。他认为, 元数据M计划的提出, 旨在优化英国的元数据市场, 以简化目录记录的供应和需求, 推动各类图书馆之间可以自由和重用书目数据。

美国国会图书馆提出的BIBFRAME基于关联数据, 在北美地区应用较广, 欧洲地区也举办了相应的研讨会。此次“BIBFRAME实践”分会场为全球各地机构与专家提供了交流空间。美国国会图书馆的Sally Mccallum介绍BIBFRAME在美国国会图书馆所面临的挑战与解决方案。她指出, 从MARC向BIBFRAME过渡时, 需要考虑关联书目数据模型的开放性与兼容性, 需要处理好书目模型的语言和底层数据脚本的支持问题。因此, 美国国会图书馆推出两个阶段性项目以实现过渡。从MARC向BIBFRAME的数据迁移还需要花费更长的时间, 人员培训、系统灵活性在BIBFRAME 100计划中面临挑战。

美国斯坦福大学图书馆的Philip E. Schreur副馆长讨论了BIBFRAME数据转换流程中的问题。面向生产的关联数据LD4P(Linked Data for Production)项目分为多个阶段进行, 包括扩大实施范围、吸引更多机构参与的LD4P2阶段, 以及面向数据转换与存储至本地系统的LD4P3阶段。

瑞典国家图书馆的Fredrik Klingwall介绍了为MARC向BIBFRAME转换数据而建设的librisXL系统, 该系统链接了多个机构资源, 提供关联数据服务和美国国会图书馆的链接。新加坡国家图书馆委员会的Linnet Ng介绍了新加坡国家图书馆基于现有资源的关联数据转化, 以及各类图书馆、档案馆资源描述的MARC格式向BIBFRAME转换的发展规划。

韩国国立图书馆的Yoon Kyung Choi介绍了主题词自动推荐, 主题词描述自动转换为KORMARC的项目做法, 从数据模式分析、系统处理设计、系统实施进行元数据流程改进等在应对韩国国家图书馆元数据工作流程中的挑战。

葡萄牙国家图书馆的Helena Patrício面对不断增长的数据资源描述需求,从数字特藏、获取服务、开放数据以及内容重用4个方面提出简化原有数据模型、使用本地标准代替部分元数据标准完善元数据模型,参考国际模型设置资源数据开放专区,以公有领域发布开放数据,每年向Europeana共享元数据。

元数据业务是信息资源建设的核心,以BIBFRAME为代表的编目工作升级,融入网络开放数据生态体现出图书馆元数据的传统编目与技术创新融合。

2.3 元数据和隐私保护

数据时代的个人信息与隐私保护越来越受到关注。元数据在很大程度上决定着隐私数据安全利用的发展走向。

爱丁堡龙比亚大学的David Haynes以美国中情局的“We kill people based on metadata”为引子,讨论类机构如何利用元数据对目标个体、社交媒体监督、基础设施进行的关联监督以及预防网络犯罪。他介绍了如何防范此类情况发生的方法,例如完善保护隐私数据的法律法规、隐私保护元数据标准的制定、相关机构对于隐私数据保护利用的政策支持以及相应的规则制定等。

都柏林圣三一大学ADAPT中心的Harshvardhan J. Pandit研究员介绍了W3C数据隐私词汇与控制社区小组(Data Privacy Vocabulary and Controls Community Group, DPVCG)提出的数据隐私词表(Data privacy vocabulary, DPV)。该词表在《欧盟通用数据保护条例》(General Data Protection Regulation, GDPR)基础上扩展,包括个人数据分类、数据收集目的分类以及处理、披露、同意利用个人数据事件的描述信息。DPV使用语义词汇术语和本体关系,表示与隐私和数据保护相关概念。

加拿大安大略省档案管理员、首席信息安全官John Roberts从自身从事政府工作的经验出发,分享其对元数据在政府政务应用与行政工作中的作用的想法。“数字安大略”的战略规划愿景在于构建一个便捷数字化、安全、互联以及数据支持的世界领先的数字管辖区。在对数据(或元数据)的管理与政策制定上,出台了一系列标准法规用于规范数据开放、存取、人工智能应用、网络安全、隐私保护、数据集成以及档案描述等,充分实现政府对元数据的管理,提高安大略省政府的行政效率。

由于元数据与隐私之间关系密切,在移动互联网、物联网、信息流推荐的数据密集型应用中,将有更多讨论。

2.4 科学数据管理

在开放科学和数字学术环境中,元数据在科学数据管理中发挥主要作用。此次年会对FAIR原则的应用、科学数据开放获取与存档等方面开展讨论。

澳大利亚研究数据共享中心的吴明芳指出建设研究资源库对于满足科研教育需求的迫切性,介绍了在构建资源库所面临的资源描述与呈现等过程中的挑战与相应的解决方案,提出遵循FAIR原则,将结构化元数据应用于资源库建设,以指南方式介绍了一系列用于网络应用和发布结构化元数据的做法。

西班牙萨拉戈萨大学的Javier Noguera Iso副教授分享了DC元数据在开放地理空间数据出版中的应用,从空间数据基础设施的背景、DC元数据作为互操作性和可查找性解决方案、开放地理空间数据时代的DC元数据以及DC元数据对于地理数据描述质量的挑战等方面进行了介绍。

加拿大西蒙菲莎大学的Emma Griffiths介绍了加拿大公共卫生综合快速传染病分析生物信息学平台IRIDA,应对食品安全保障和基因流行病学研究,在基因本体基础上开发了Foodon和GenEpio两个本体,实现了本体搜索OLS、自动标引LexMapr、基因跟踪GenomeTrakr等功能。

在促进食品农业科学数据获取方面,联合国粮农组织(Food and Agriculture Organization of the United Nations, FAO)高级信息管理官Imma Subirats Coll介绍了FAO的AGRIS现状,当前已有454个数据提供者向FAO提交了数据集,以促进科学数据共享,分享了AGRIS多语种支持,以及向AGRIS提交元数据的方式。

2.5 元数据技术应用

DCMI年会的传统是突出实务性,始终关注元数据相关技术工具和系统应用。此次年会设有一个前沿话题“Why AI≠Automated Indexing”,从人工智能技术赋能元数据角度展开探讨。

荷兰历史学家Hans Brandhorst介绍了其创建的Iconclass系统,这是全球最大的基于杜威十进制分类

法的视觉艺术分类系统,用于对各类历史图片和现代摄影照片进行分类与描述。2015年该系统网站以关联数据形式对外开放,2021年网站更新,进一步提升用户体验。最新的Iconclass系统支持多语种,如葡萄牙语、荷兰语以及部分中文等,还提供交流论坛、机构API接入以及书目数据分类等功能。面向图片分类识别的更新,通过自动图片索引方式区分不同图片含义,在测试大量的图片集用于机器学习后,系统能更加精确面向图片进行细节描述分类与自动标引。

西班牙巴塞罗那超级计算机中心的Joaquim More López等以“Saint George On a Bike”为主题,说明人工智能在图片识别和自动图片索引的应用。巴塞罗那超级计算机中心的Marinescu Maria-Cristina指出,将人工智能用于图片识别的最初动机是元数据的自动标注、用户通过网页或应用程序交互,提高搜索引擎检索效率以及提供用户无障碍访问等。他认为,人工智能逐渐应用于文化遗产领域的图像自动索引,将极大提高不同时空范围的图片内容描述效率。以Saint George on a Bike数据库为例,该数据库包含15000多幅图像数据,提供API允许第三方采集图像数据。将该数据集作为图像识别训练集,根据算法识别出不同图像描述语句的主语、谓语与宾语,从而实现不同图像之间视觉关系的精确分类索引。

美国资深词表专家Marjorie Hlava结合文献研究和自身词表开发经验,对自动标引和人工智能中两个“AI”谈及个人理解。芬兰国家图书馆的Osma Suominen介绍了应用于芬兰文化与科学资料库Finna的自动主题索引和分类的工具Annif以及关于该工具的开发过程、开发目标、使用的目标索引质量、参与的社区建设和开发工具的安装部署流程。

除此之外,此次年会的技术工具培训与系统使用分享也非常丰富。美国密歇根州立大学的Sharon Leon副教授介绍了面向文化遗产资源研究、发布关联数据和资源管理的平台Omeka。Omeka是一个由多个组织赞助支持、服务于文化遗产研究的数字学术项目。2016年推出新一代的Omeka S,在Omeka Classic开源基础上做出大量优化,实现多项目、跨站点资源聚合管理,实现了对单一资源对象实体的数字化展览与保管,以JSON-LD数据格式发布并对外提供API。Omeka S能够从其他机构知识库系统Fedora和DSpace导入集成数据。

Coli-Conc开源系统由德国公共图书馆网络总部开发,用以促进图书馆知识组织系统之间的索引管理和

交换。该系统提供基于网络访问的Cocoda工具,为用户知识组织系统的标引与映射编辑功能,以可视化浏览方式展示知识组织系统,支持创建和修改映射候选术语并分配映射类型,并对外提供开源,后端服务则包括术语发布与检索服务、映射与存储、映射建议与映射质量统计等服务。

出版供应商Casalini Libri首席信息官Tiziana Possemato介绍了面向文化机构的关联数据管理系统Share-VDE。该系统致力于为图书馆和各类文化数据机构提供一套数据的关联管理方案,使用关联数据技术进行关联和呈现日益增长的海量数据,促进书目数据的交换重用,提高知识可视化与知识获取的便捷性。匈牙利塞切尼国家图书馆的Miklós Lendvay介绍了匈牙利国家图书馆项目平台HNLP,该平台旨在整合多源异构数据,以分布式、模块化促进数字资源共享与协作。德国斯图加特传媒学院Kai Eckert教授介绍了以关联数据形式发布犹太文化,分享了历史资源平台JudaicaLink的实体管理过程等。

3 总结与展望

DC元数据从初始制定到推广应用历经了26年,举办了29届年会。发展至今,DCMI年会是了解元数据最新研究与应用的学术活动,其讨论与交流内容反映了当下元数据研究面临的热点问题与发展态势。

2020年是DCMI成立25周年。在回顾和展望未来的当下,DCMI 2020年会主旨报告第一位专家是DCMI组织的创始人之一Stuart Weibel,作题为“2020 Vision: Reflections on a Quarter Century of Metadata”的主旨报告^[1]。这份主旨报告串起了元数据的过去、现在与未来,引发元数据研究者、实践者及教育者的集体共鸣。在25周年节点上,DCMI 2021虚拟年会承上启下,继往开来,办成了至今为止最盛大的一次元数据学术盛会。此次年会带来丰富的元数据研究素材与实践经验,为国内元数据相关研究者、从业者以及高校专业师生开拓了视野,启发了思路。元数据作为数据基础设施建设的核心要件,相关的技术要素活跃,语义增强与关联集成是当前元数据助推行业 and 主题数据资源建设的核心任务。面向数字人文的文化遗产元数据、隐私保护、科学数据管理等是元数据应用领域的新开拓。

随着DC元数据的全球推广,国内对DC元数据有着持续关注和本地化应用。2004年国内图书情报机构上

海图书馆承办了DC-2004年会。2006—2013年,国内专家学者曾对DCMI年会内容持续跟踪与评述。近些年,尽管DCMI年会都有国内专家学者参会,但有关的年会进展与评述鲜见于公开发表,造成跟踪中断。对DCMI年会的持续关注、积极参与、国际交流及评述转化,是促进国内元数据研究与应用保持活力的一种推动力。

国内在元数据标准制定、各类信息资源建设中已取得阶段性进展,学界与业界对语义网、关联数据、本体的技术理念理解与应用手段也已达到成熟阶段,但依然缺乏像芬兰语义门户、Europeana等具有影响力的本土项目。在数智赋能的数字文明新时代,中国元数据研究与实践应守好数据基础设施建设盘面,加强与国外交流合作,在数据新基建中,以元数据为助推器,以智慧数据作为目标,用好数据,开拓创新多元化应用与服务。

参考文献

- [1] DCMI Virtual 2021 [EB/OL]. [2021-11-17]. <https://www.dublincore.org/conferences/2021/>.
- [2] HYVÖNEN E. “Sampo” Model and Semantic Portals for Digital Humanities on the Semantic Web [C] //In Proceedings of the Digital Humanities in the Nordic Countries 5th Conference (DHN). 2020: 373-378.
- [3] Sampo Model and Series of Semantic Portals [EB/OL]. [2022-04-01]. <https://seco.cs.aalto.fi/applications/sampo/>.
- [4] National Semantic Web Ontology Project in Finland (FinnONTO), 2003-2012 [EB/OL]. [2022-03-01]. <https://seco.cs.aalto.fi/projects/finnonto/>.
- [5] HYVÖNEN E, TUOMINEN J, ALONEN M, et al. Linked Data Finland: A 7-star model and platform for publishing and re-using linked datasets [C] //European Semantic Web Conference. Cham: Springer, 2014: 226-230.
- [6] HYVÖNEN E. Using the Semantic Web in Digital Humanities: Shift from data publishing to data-analysis and serendipitous knowledge discovery [J]. Semantic Web, 2020, 11 (1): 187-193.
- [7] BRICKLEY D. What is left to do? [EB/OL]. [2022-03-01]. <https://dcevents.dublincore.org/IntConf/index/pages/view/speakers-2012.html>.
- [8] Understand how structured data works [EB/OL]. [2021-11-17]. <https://developers.google.com/search/docs/advanced/structured-data/intro-structured-data#structured-data>.
- [9] WIJESUNDARA C, SUGIMOTO S. Metadata model for organizing digital archives of tangible and intangible cultural heritage, and linking cultural heritage information in digital space [J]. LIBRES: Library and Information Science Research Electronic Journal, 2018, 28 (2): 58-80.
- [10] MIHARA T. Designing the Metadata Model for the Aggregation of MAG Metadata-the Development of Media Art Database [EB/OL]. [2021-11-17]. https://www.dublincore.org/conferences/2021/presentations/cultural_metadata_for_what_entities_are_we_creating_metadata/slides/tetsuya_mihara.pdf.
- [11] Stuart Weibel. 2020 Vision: Reflections on a Quarter Century of Metadata [EB/OL]. [2021-11-17]. https://www.dublincore.org/conferences/2020/presentations/2020_vision_reflections_on_a_quarter_century_of_metadata/.

作者简介

范炜,男,1981年生,博士,副教授,研究方向:信息组织与信息检索,E-mail: fanw@scu.edu.cn。

林君松,男,1996年生,硕士研究生,研究方向:信息检索与情报服务。

Overview of DCMI 2021 Virtual Conference

FAN Wei LIN JunSong

(School of Public Administration, Sichuan University, Chengdu 610064, P. R. China)

Abstract: DCMI annual conference is an important international academic conference in metadata field. The article summaries the research topics, project experiences and technology implementation for DCMI 2021 virtual conference. This conference devoted continuous attention to enable semantic web, linked data and ontology technology for metadata practices, shared the innovative ideas and experiences of Finnish national ontologies and semantic portals, and discussed the integrated development of DC metadata and Schema.org. In metadata applications, it mainly involves cultural heritage metadata, bibliographic resource conversion and uprate, privacy protection, research data management, etc. In metadata technology systems, it discussed artificial intelligence and automatic indexing, entity identification and management, linked data publish and digital resource management platform, etc. Track the latest progress of DCMI conference, it provides valuable references for China's local metadata research and application practices.

Keywords: DCMI; 2021 Virtual Conference; Metadata Innovation; Semantic Web; Linked Data; Ontology; Diversity

(收稿日期: 2022-03-05)