

德国政府开放数据的元数据标准及应用^{*}

翟军 毕智航 董可新 沈立新
(大连海事大学航运经济与管理学院, 大连 116026)

摘要: 德国致力于成为世界上领先的开放数据国家, 通过开放数据促进创新发展, 尤其是在2021年密集出台了开放数据的战略、行动计划和法律, 促使元数据标准建设呈现出新的特点和发展趋势。采用网络调查和文献分析方法, 介绍德国元数据标准DCAT-AP.de的主要内容和应用情况, 提炼可供借鉴的经验和做法。对我国而言, 应在元数据标准的政策和法律支持、数据治理和组织保障、本土化与国际化的有机结合等方面借鉴其先进经验。

关键词: 开放政府数据; 元数据标准; 数据目录; 数据治理; DCAT

中图分类号: G254 **DOI:** 10.3772/j.issn.1673-2286.2022.12.001

引文格式: 翟军, 毕智航, 董可新, 等. 德国政府开放数据的元数据标准及应用[J]. 数字图书馆论坛, 2022(12): 1-9.

数字经济时代, 数据的开放共享有效推动着数据要素的价值释放^[1-2]。德国的数字经济规模居世界第三位, 仅次于美国和中国, 2020年达到2.54万亿美元^[3], 其中由开放数据带来的效益日益显著^[4]。根据知名的康拉德·阿登纳基金会(Konrad-Adenauer-Stiftung)的研究, 2016年开放政府数据(Open Government Data, OGD)为德国贡献了431亿欧元的经济附加值, 到2025年这一数字有望增加到1 311亿欧元^[5]。

致力于成为世界领先的开放数据国家^[6], 德国政府于2013年签署《G8开放数据宪章》(G8 Open Data Charter)^[7], 2016年加入开放政府合作组织(Open Government Partnership, OGP), 2021年采纳《国际开放数据宪章》(International Open Data Charter)^[8]。在77个OGP成员中, 德国是率先提交《2021—2023年国家行动计划》(National Action Plan)的国家之一^[9]。2021年7月, 德国开始实施为期5年的《联邦政府开放数据战略》(Open-Data-Strategie der Bundesregierung)^[10]。在法律建设上, 2017年7月, 德国《电子政务法》(E-Government-Gesetzes, EGovG)修正案新增了开放数据的条款(第

12a节), 也被称为《第一部开放数据法》(Ersten Open-Data-Gesetz)^[11]。2021年6月, 德国联邦议院通过《第二部开放数据法》(Zweites Open-Data-Gesetz, 即《电子政务法》新的修正案)和《数据使用法》(Daten-Nutzungs-Gesetz)^[9], 进一步夯实了开放数据的法律基础。

经过不懈的努力, 德国的OGD有了长足的发展。根据国际组织开放数据观察(Open Data Watch)的2020/21 ODIN(Open Data Inventory)评估报告, 德国的ODIN得分是77分(满分100分), 全球排名第13位^[12]。在欧盟的开放数据成熟度(Open Data Maturity)评估中, 德国的排名从2016年的第21位上升到2020年的第8位, 政策/法律和网站/元数据两个维度的得分分别是97%和90%, 远高于平均水平(85%和79%)^[13]。

作为重要的数据基础设施(data infrastructure), 元数据标准(metadata standards)得到了美国^[14]、英国^[15]、澳大利亚^[16]和欧盟^[17]的普遍重视。同样地, 元数据标准建设也是德国实施开放数据法律和战略的主要内容, 并呈现新的特点和发展趋势^[18], 对德国政府开放数

^{*} 本研究得到辽宁省普通高等教育本科教学改革研究项目“数智时代信息管理与信息系统一流专业建设的研究与实践”(编号: 辽教通[2022]166号-339)资助。

据的元数据标准建设进行系统分析可为我国各级政府的开放数据实践及其元数据建设提供借鉴和参考。

1 元数据标准

德国《开放数据法》规定(见EGovG 12a节第5条),政府数据资源的元数据应遵循统一的标准,并发布到国家元数据平台GovData(www.govdata.de)^[19]。2017年6月,在W3C数据目录词汇表(Data Catalog Vocabulary, DCAT)和欧盟DCAT应用纲要(DCAT Application Profile, DCAT-AP)的基础上,jinit[AG公司研制出元数据方案DCAT-AP.de,并转让给GovData团队。2018年6月,德国IT规划委员会(IT-Planungsrat)将DCAT-AP.de确立为正式强制性标准^[20]。2021年7月,《联邦政府开放数据战略》指出,要进一步完善DCAT-AP.de标准,使其适应新的技术发展,如实时数据的开放需求等^[10]。2022年3月1日,DCAT-AP.de V2.0版本发布^[21],成为W3C DCAT标准谱系中的新成员,谱系的其他成员还包括美国的DCAT-US v1.1、意大利的DCAT-AP_IT、荷兰的DCAT-AP-NL及欧盟的DCAT-AP等。

作为一种实用且完备的元数据标准,DCAT-AP.de由三部分组成:模型与元素定义;受控词汇表与URI(Uniform Resource Identifier)设计;技术与使用指南。

1.1 元数据模型

为提高元数据的互操作性和一致性,GovData团队采用元数据应用纲要(Metadata Application Profiles, MAP)的方式设计DCAT-AP.de的本体模型,集成和复用的主要词汇表(本体)见表1^[21],包括被搜索引擎公司支持的Schema.org词汇表。

表1 DCAT-AP.de复用的主要词汇表

序号	词汇表的命名空间	前缀	说明
1	http://www.w3.org/ns/dcat#	dcat	W3C DCAT词汇表
2	http://data.europa.eu/r5r/	dcatap	欧盟DCAT-AP
3	http://purl.org/dc/terms/	dct	都柏林核心元数据
4	http://www.w3.org/ns/adms#	adms	数字资产元数据
5	http://xmlns.com/foaf/0.1/	foaf	FOAF词汇表
6	http://schema.org/	schema	Schema.org词汇表
7	http://www.w3.org/2004/02/skos/core#	skos	W3C SKOS词汇表

DCAT-AP.de模型包含25个类/实体,分为5个强制类、4个推荐类和16个可选类,一些主要的类及其属性见表2,保持了与欧盟DCAT-AP的语义一致性^[17]。同时,为了满足数据资源的版权管理、质量管理、可用性管理和行政区域管理等本土化需求,相对于欧盟DCAT-AP,DCAT-AP.de定义了一组新的元素/属性(见表3),它们的命名空间是<http://dcat-ap.de/def/dcatde/>,前缀是dcatde。

表2 DCAT-AP.de的主要类及其属性

序号	类的德文名称	类的URI	主要属性	说明
1	Katalog	dcat:Catalog	dcat:dataset, dct:description, dct:publisher, dct:title, foaf:homepage, dcat:themeTaxonomy	数据目录
2	Datenstruktur	dcat:Dataset	dct:description, dct:title, dcat:distribution, dcat:keyword, dct:publisher, dcat:theme, dcatde:politicalGeocodingURI等	数据集
3	Distribution	dcat:Distribution	dct:title, dct:format, dct:license, dct:modified, dcat:accessURL, dcatde:plannedAvailability等	数据资源
4	Verantwortliche Stelle	foaf:Agent	foaf:name, dct:type	机构
5	Lizenzdokument	dct:LicenseDocument	dct:type	法律文件
6	Kategorie	skos:Concept	skos:prefLabel	分类/主题

1.2 受控词汇表

为了实现与欧盟DCAT-AP的相互兼容及规范元素的

取值,DCAT-AP.de引入了较多的受控词汇表(Controlled Vocabularies)。受控词汇表分成两类:①来自欧盟的有11个,如数据主题词汇表、语言词汇表和文件格式词汇表

等^[17]; ②自定义的有12个(见表4), 它们不仅支持德国的行政区编码等编码方案, 也支持统计数据和时间序列数据的开放, 以满足本国的多方面需求。

表3 DCAT-AP.de新定义的本土元素/属性

序 号	元素/属性名称	定义域	值 域	说 明
1	dcate:maintainer	dcate:Dataset	foaf:Agent	数据集的维护者
2	dcate:originator	dcate:Dataset	foaf:Agent	版权所有人
3	dcate:contributorID	dcate:Dataset	rdfs:Resource	数据提供方的ID
4	dcate:politicalGeocodingURI	dcate:Dataset	rdfs:Resource	数据覆盖的行政区的URI编码
5	dcate:politicalGeocodingLevelURI	dcate:Dataset	rdfs:Resource	数据覆盖的行政区级别, 如联邦、州等
6	dcate:qualityProcessURI	dcate:Dataset	rdfs:Resource	质量保障过程
7	dcate:geocodingDescription	dcate:Dataset	rdfs:Literal	数据覆盖的地理范围
8	dcate:legalBasis	dcate:Dataset	rdfs:Literal	法律依据, 如《信息自由法》等
9	dcate:plannedAvailability	dcate:Distribution	rdfs:Resource	数据资源的可用性
10	dcate:licenseAttributionByText	dcate:Distribution	rdfs:Literal	许可的详细信息

表4 DCAT-AP.de自定义的受控词汇表(部分)

序 号	词汇表名称	词汇表的URI	说 明
1	数据提供方	http://dcat-ap.de/def/contributors	含有59个词汇, 每个词汇为一个数据提供方的URI
2	数据集类型	http://dcat-ap.de/def/datasetTypes	仅含有1个词汇, 表示时间序列类型
3	开放许可	http://dcat-ap.de/def/licenses	含有34个词汇, 表示可供选择的开放许可列表
4	可用性级别	http://dcat-ap.de/def/plannedAvailability	含有4个词汇: 暂时可用、一年内可用、几年内可用和长期可用
5	行政区级别	http://dcat-ap.de/def/politicalGeocoding/Level	含有6个词汇, 分别是国际、欧盟、联邦、联邦州、行政区和市
6	联邦州编码	http://dcat-ap.de/def/politicalGeocoding/stateKey	含有16个词汇, 每个词汇对应一个联邦州的URI编码
7	地区编码	http://dcat-ap.de/def/politicalGeocoding/districtKey	含有617个词汇, 每个词汇对应一个地区的URI编码

GovData团队采用W3C SKOS标准和RDF语法等语义Web技术定义这些词汇表, 每个词汇表及其中的词汇都被分配了一个持久和可靠的URI, 一些例子见表4和表5。为此, DCAT-AP.de标准专门提供URI设计公约规范URI的形式和使用, 主要包括: ①URI的模板为http://dcat-ap.de/def/{konzept}; ②URI中的“概念”(konzept)使用英语表达; ③从2017年7月1日开始, URI的有效期至少为10年。

这些词汇表一经定义就保持稳定, 如有变化则反映在新的版本中。例如, 数据提供方词汇表从2018年至今已有14个版本, 最新的版本是2022年6月23日发布的。

在开发元数据标准时, 大量使用语义Web受控词汇

表, 以描述本土需求和变化, 代表着DCAT元数据规范的发展方向。

1.3 技术与使用指南

不同于W3C DCAT和欧盟DCAT-AP这些高层次的指导类规范, DCAT-AP.de标准进一步通过《技术与使用指南》(以下简称《指南》)指导实际工作中元数据记录的编写和生成。《指南》共有41条, 值得关注的有以下4点: ①加强对元数据的语义约束, 建立了元素/属性与受控词汇表的对应关系(见表6), 明确这些元素的取值必须来自欧盟或德国的受控词汇表; ②虽

然在受控词汇表中有34个开放许可可供选择,但《指南》将dl-zero-de/2.0(德国数据许可)和cc-by/4.0(国际知识共享许可)确立为推荐的开放许可,建议优先使用这两个许可;③规定采用开放标准编码元数据记录(语法约束),包括W3C的RDF/XML和RDF-Turtle标准,在V1.1版本中增加了JSON-LD(JSON for Linked

Data)编码格式;④规定验证机制,即对元数据记录是否符合元数据标准的模式和约束进行定义,在V1.1和V2.0版本中使用W3C SHACL(Shapes Constraint Language)验证方案,以前的版本则使用W3C XSD(XML Schema Definition)验证方案,这在各国(包括欧盟)的元数据标准建设中是超前的。

表5 “可用性级别”受控词汇表中的词汇

序 号	词汇的URI	说 明
1	http://dcat-ap.de/def/plannedAvailability/temporary	数据暂时可用,随时可能消失
2	http://dcat-ap.de/def/plannedAvailability/experimental	数据通过实验获得,一年内可用
3	http://dcat-ap.de/def/plannedAvailability/available	几年内可用,可用于中期规划
4	http://dcat-ap.de/def/plannedAvailability/stable	数据长期可用

表6 元素/属性与受控词汇表的对应关系

序 号	元素/属性	受控词汇表	取值示例
1	dcat:theme (数据集的主题)	数据主题 (欧盟)	http://publications.europa.eu/resource/authority/data-theme/GOVE
2	dct:type (数据集的类型)	数据集类型 (德国)	http://dcatap.de/def/datasetTypes/collection
3	dcatde:contributorID (数据提供方)	数据提供方 (德国)	http://dcat-ap.de/def/contributors/berlinOpenData
4	dcatde:plannedAvailability (数据资源的可用性)	可用性级别 (德国)	http://dcat-ap.de/def/plannedAvailability/stable
5	dct:license (数据资源的许可)	开放许可 (德国)	http://dcat-ap.de/def/licenses/cc-by/4.0
6	dct:language (数据资源的语言)	语言 (欧盟)	https://publications.europa.eu/resource/authority/language/DEU
7	dct:format (数据资源的格式)	文件格式 (欧盟)	https://publications.europa.eu/resource/authority/file-type/CSV

2 元数据标准的应用

德国是具有分权传统的联邦制国家,采取三级公共行政管理层级。联邦下辖16个具有较高自治权的联邦州,州管辖市/县级地区。各个州和地方政府在OGD的政策和实践上具有一定的自主性^[9]。为协调联邦政府与地方政府的工作,德国采用了中央和地方共治共管的开放数据治理模式^[13],成功做法之一是“数据分散存储、元数据集中管理”。为实现这一点,元数据标准DCAT-AP.de的推广应用起到了关键作用。

2.1 数据目录间的元数据采集和交换

2015年1月,德国国家数据门户网站GovData.de正式上线运行,成为联邦政府和地方政府发布数据的统一平台。一些地方政府也陆续上线了开放数据网站,目前总数达近百个。德国的开放数据网站主要采用CKAN(Comprehensive Knowledge Archive Network)开源平台进行数据集的元数据管理,即数据目录(Data Catalogue)管理^[22]。除了通用的开放数据网站,在欧盟空间信息基础设施(Infrastructure for Spatial Information in the European Community, Inspire)计划框架下,德国

还建有专用的地理数据平台, 如www.geoportal.de和www.geoportal.nrw等。

德国在IT规划委员会的统一规划下, 采取了“自上而下”的元数据标准推广方式。从2019年起, GovData.de全面采纳DCAT-AP.de标准; 随后, DCAT-AP.de逐步推广到各地方政府的开放数据网站。例如, 2021年7月14, 波恩市完成了旧的元数据到DCAT-AP.de的转换^[23]。借助于统一的DCAT-AP.de标准, 各个数据目录间实现了元数据的自动采集和交换, 保证元数据记录一次发布, 就能第一时间出现在各级目录网站中。例如, 波恩市的(交通和道路)网络拓扑图(Netztopologie Stadtplan Bonn)数据集的元数据以机器可读的格式(JSON-LD等)和面向人的阅读格式(HTML网页)发布到开放数据网站(opendata.bonn.de), 随后被自动采集到上一级的北莱茵-威斯特法伦州(Nordrhein-Westfalen, 北威州)的网站(open.nrw), 继而同步出现在国家数据平台(GovData.de)上。最后, 通过GovData.de, 该数据集的元数据出现在欧盟数据门户(data.europa.eu)上, 被更广泛的用户查询、下载和使用。

这样, 在欧盟环境下, 德国的OGD就形成了四层数据目录聚合机制, 最基层是市/县政府的开放数据网站, 中间两层(核心层)是各州和国家的数据平台, 最高层是欧盟的统一数据门户网站。截至2022年11月底, 除了萨克森-安哈尔特和萨尔州, GovData.de已经能够采集14个州的元数据, 涉及数据集7万多个。在联邦州的层面, 北威州的数据平台聚集了近50个外部数据目录的元数据。在最高的欧盟层级, 欧盟数据平台聚集了36个国家的176个数据目录, 其中3个来自德国, 分别是通用数据目录平台GovData、地理数据目录平台GDI-DE和开源地理软件平台52° North。在这个过程中, 德国的元数据标准DCAT-AP.de与国家数据网站GovData.de共同起到了承上启下的中介协调和质量保障作用; 而欧盟的元数据标准DCAT-AP和欧盟数据门户网站则发挥着规范引领和质量监测作用。

2.2 元数据采集器和API

目录聚合(catalogues aggregation)是上一级的数据目录通过采集器(harvester)软件模块调用下一级数据目录的元数据API自动实现的。为此, CKAN平台扩展了新的插件——DCAT-AP.de采集器, 专门采集RDF和JSON格式的DCAT-AP.de元数据。同时, CKAN

平台提供DCAT端点(Endpoint, API的一种), 向外暴露RDF/XML、RDF-Turtle和JSON-LD格式的元数据文件, 分为数据集和数据目录两种粒度。这需要建立DCAT元数据与CKAN元数据的映射(见表7), 以实现两者之间的相互转换^[24]。

表7 DCAT元数据与CKAN元数据的映射(部分)

序 号	DCAT属性	CKAN字段
1	dct:title	title
2	dct:description	notes
3	dcat:keyword	tags
4	dcat:landingPage	url
5	foaf:page	extra:documentation

GovData和北威州的数据平台还提供了CSW(Catalogue Service for the Web)采集器和OAI-PMH(Open Archives Initiative Protocol for Metadata Harvesting)接口, 实现ISO 19139元数据与DCAT-AP.de元数据的映射与转换, 以适应地理数据平台与通用数据平台之间的元数据交换。

由于大量采用语义Web技术, GovData平台还提供SPARQL查询端点(API的一种), 供客户端对元数据进行更加灵活和复杂的查询。例如, 下面的语句查询来自北威州的数据集个数, 返回结果如下。

```
"6452"^^<http://www.w3.org/2001/XMLSchema#integer>
PREFIX dcatde: <http://dcat-ap.de/def/dcatde/>
SELECT (COUNT(?sub) AS ?datasets) WHERE {
    ?sub dcatde:contributorID <http://dcat-ap.de/def/contributors/openNRW> .}
```

其中, 三元组模式的谓词dcatde:contributorID是DCAT-AP.de定义的新属性(见表3), 其值来自数据提供方受控词汇表(见表4和表6)。

2.3 元数据验证与质量保障

元数据的自动采集与发布, 对元数据的质量提出了更高的要求。为保障各个数据提供者提供的元数据记录符合DCAT-AP.de所定义的模式和约束, GovData平台联合欧盟互操作性测试平台ITB(Interoperability Test Bed)推出数据验证(Data Validation)服务——DCAT-AP.de验证器(DCAT-AP.de Validator)^[18], 这在

全球范围尚属首次。该验证器于2020年3月12日首次上线,采用W3C SHACL验证机制,可以验证3个级别,分别是:①欧盟DCAT-AP符合性验证;②德国DCAT-AP.de符合性验证;③GovData平台验证。验证的内容主要有:①数据集的元数据记录是否含有强制属性和推荐属性;②属性是否满足基数约束;③属性取值是否来自受控词汇表。

在DCAT-AP.de验证器的基础上, GovData平台为注册用户元数据质量仪表盘服务,监测元数据对DCAT-AP.de标准的遵守程度及元数据文件链接的可用

性等指标,优化了元数据的采集和发布流程,见图1^[18]。这一做法已经被欧盟开放数据团队推荐给欧盟各成员国^[13]。2022年5月,北威州数据平台也上线了元数据质量仪表盘模块(mqa2-open.nrw.de),从6个维度(可发现、可获取、互操作、可重用、背景描述和综合指标)监测各个数据目录的元数据质量。2022年11月9日,重新设计的欧盟数据平台(data.europa.eu)上线^[25],改进之一是在每个数据集页面上增加了元数据质量监测栏目,给出各个质量维度的得分。

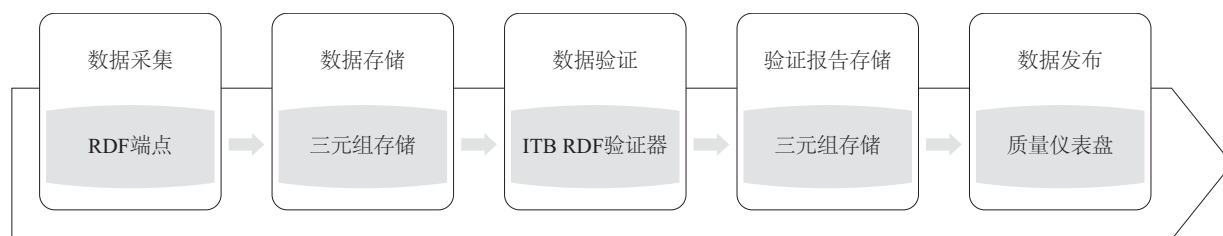


图1 GovData平台元数据质量仪表盘的工作流程

3 德国元数据标准建设的经验与启示

元数据标准建设与数据目录编制是我国开放数据的基础及核心工作之一^[26-27]。广东、山东、贵州、内蒙古和四川等省/自治区率先研制和发布了开放数据元数据的地方标准。2019年以来,上海、贵州、浙江、山东和重庆等省/直辖市政府或人大陆续出台公共数据开放办法或条例,确立了“需求导向、有序开放、安全可控、统一标准”的工作原则,规定公共数据开放要实行统一的目录管理。2022年6月,国务院印发《关于加强数字政府建设的指导意见》,在“构建开放共享的数据资源体系”部分强调,编制公共数据开放目录及相关责任清单,构建统一规范、互联互通、安全可控的国家公共数据开放平台^[28]。但实践中,我国的OGD存在元数据规范和标准不统一、数据分类方式不一致、数据目录编制主体分散、元数据质量有待提高等问题和挑战^[29],还需要积极借鉴德国的先进经验。

3.1 通过开放数据行动战略和法律,全方位推进元数据标准的建设和实施

为贯彻《G8开放数据宪章》,德国于2014年11月开

始实施《G8开放数据宪章国家行动计划》,明确在建设国家数据平台GovData的同时,必须实现元数据的标准化,确保元数据的质量和互操作。接下来,在第一轮《OGP国家行动计划》(2017—2019年)中承诺,由联邦政府IT规划委员会负责颁布和实施开放数据的元数据标准。2021年6月通过的《第二部开放数据法》,确立了GovData作为全国统一的元数据平台的法律地位^[11]。

2021年1月,德国开始实施《联邦政府数据战略(2021—2025)》^[30],在高价值公共数据开放领域引入FAIR (Findable, Accessible, Interoperable, Reusable)原则,对元数据标准的互操作性提出了更高的要求。2021年7月,《联邦政府开放数据战略》进一步明确完善DCAT-AP.de、保障元数据质量的负责部门和工作任务^[10]。

不仅德国,美国、英国等OGD先进国家都将元数据标准建设纳入国家战略高度,及时开发、升级和推广应用统一的元数据标准。我国数字政府建设和全国一体化政务大数据体系建设中,也需要国家层面的统一元数据标准,以实现全国范围内数据目录和标准规范的一体化^[28,31]。但我国还没有开放数据元数据的国家标准,各地方标准间存在异构问题,导致互操作性差,阻碍了统一的国家数据开放平台和流通市场的形成。

3.2 加强元数据治理, 完善元数据标准施行的组织保障

DCAT-AP.de是德国的强制性元数据标准, 将其应用到近百个地方政府的开放数据网站上, 是一项复杂而持久的工作。为此, 正在实施的《开放数据战略》和第三轮《OGP国家行动计划》(2021—2023年)将地方政府的数据开放工作纳入国家行动计划, 同时指定具体机构负责中央和地方的协调工作。

德国公共行政数字化的中央管理机构IT规划委员会负责协调联邦和州政府在信息技术上的合作, 解决元数据标准化和软件体系结构等问题。联邦行政办公室(BVA)下属的开放数据能力中心(Competence Center for Open Data)提供技术方案, 连接中央和地方的开放数据平台, 实现元数据的采集和聚合。这些OGD混合治理模式在组织机构上的制度安排, 保障了元数据标准在德国全国范围内的贯彻施行。

在我国, 元数据的地方标准尚没有充分应用到各地、各级的开放数据网站中, 更缺乏元数据的验证和质量保障机制。例如, 广东省和深圳市的开放数据网站(gddata.gd.gov.cn, opendata.sz.gov.cn)实际使用的元数据在元素名称、取值和个数上存在不一致现象, 贵州省(data.guizhou.gov.cn)和贵阳市(data.guiyang.gov.cn)也存在类似问题。解决这一问题的首要工作就是要加强政府数据治理和省市协调, 明确施行元数据标准的责任机构和工作任务。

3.3 元数据建设本土化与国际化的有机结合

德国的数字市场是欧盟数字单一市场的有机组成部分。在法律制定上, 德国的开放数据法受到欧盟《开放数据和公共部门信息再利用指令》(Directive on Open Data and the Re-use of Public Sector Information)和《通用数据保护条例》(General Data Protection Regulation)的影响。相应地, 在元数据标准的建设上, 德国做到了本土化与国际化的有机结合。DCAT-AP.de是W3C DCAT和欧盟DCAT-AP的“应用纲要”, 在保持与它们互操作的基础上, 进行了适当的本地化扩展。同时, DCAT-AP.de标准的各个部分都采用了开放标准和语义Web技术, 实现了最大程度的国际化。

元数据标准本土化与国际化的统一, 促进了德国的

开放数据资源的跨境流动, 也使其成为国际数据生态系统的重要一员。2020年1月, 谷歌正式上线专业搜索引擎——数据集搜索(Dataset Search), 收录了全球数千个网站中的约2 500万个数据集^[32], 为国际数据生态系统的形成创造了基本环境。借助W3C DCAT, 可以很容易地将DCAT-AP.de映射到谷歌所支持的Schema.org数据集词汇表, 为谷歌数据集搜索服务收录、索引和查询DCAT-AP.de元数据记录铺平了道路。

在应对新冠肺炎疫情的国际合作中, 构建开放可信的国际数据生态系统日益得到各国的重视^[33]。我国的开放数据, 包括新冠肺炎疫情数据, 也是国际数据生态系统的重要一环。但我国现行的开放数据元数据标准和各网站实际使用的元数据与国际标准(W3C DCAT和Schema.org等)和技术(主要是语义Web技术)不兼容, 导致通过搜索引擎(如百度、必应和谷歌等)难以细粒度地发现和定位到一个开放数据集, 不利于数据的大范围流动和价值生成。

依托语义Web技术所形成的W3C开放标准, 在推动开放数据的语义互操作和国际化上正发挥着积极作用^[34-36]。因此, 可以借鉴德国的经验, 在我国统一的元数据标准规范中引入W3C DCAT数据模型和核心元素, 并结合我国实际, 采用W3C SKOS标准构建“数据主题”“更新频率”“开放方式”“开放类型”和“文件格式”等受控词汇表, 以满足大范围互操作、数据集搜索和本土化集成的需求。

4 结语

相对于美国、英国等开放数据先驱国家, 德国的开放政府数据行动起步较晚, 但德国能够正视差距, 以“工匠精神”稳步推进相关工作。进入2021年, 德国密集出台了开放数据的战略、行动计划和法律, 将通过开放数据促进创新发展推上了一个新高度, 展现出强劲的后发优势。本文聚焦新形势下德国的元数据标准建设, 特别是作为国际主流的W3C DCAT元数据标准谱系中重要且崭新一员的DCAT-AP.de, 介绍了它的主要内容、实施情况和保障措施, 总结了可以借鉴的特点、经验。后续研究将继续关注德国开放数据战略框架下元数据标准在实时数据、科学数据和地理数据等领域的扩展和应用情况, 为我国开放数据的标准建设和数据治理提供更多有价值的建议和参考。

参考文献

- [1] 徐绪堪, 李一铭, 庞庆华. 数字经济下政府开放数据共享的演化博弈分析[J]. 情报杂志, 2020, 39(12): 87, 119-125.
- [2] 中国数字经济发展白皮书[EB/OL]. [2022-10-18]. http://www.caict.ac.cn/kxyj/qwfb/bps/202104/t20210423_374626.htm.
- [3] 全球数字经济白皮书——疫情冲击下的复苏新曙光[EB/OL]. [2022-10-18]. http://www.caict.ac.cn/kxyj/qwfb/bps/202108/t20210802_381484.htm.
- [4] Open Data: Mit öffentlichen Daten digitale Wirtschaft fördern[EB/OL]. [2022-11-12]. <https://www.bmwk.de/Redaktion/DE/Artikel/Digitale-Welt/open-data.html>.
- [5] DAPP M M, BALTA D, PALMETSHOFER M W, et al. Open Data: the Benefits[EB/OL]. [2022-11-12]. <https://www.kas.de/de/einzeltitel/-/content/open-data.-the-benefits1>.
- [6] 赵锐. 德国: 将工匠精神融入开放数据国家行动计划[J]. 中国信息界, 2019(1): 80-84.
- [7] 陈美. 德国政府开放数据分析及其对我国的启示[J]. 图书馆, 2019(1): 52-57, 94.
- [8] Government Adopters[EB/OL]. [2022-10-22]. <https://opendatacharter.net/government-adopters/>.
- [9] Germany has submitted an action plan for the 2021-2023 cycle[EB/OL]. [2022-10-24]. <https://www.opengovpartnership.org/members/germany/>.
- [10] Open-Data-Strategie der Bundesregierung[EB/OL]. [2022-10-24]. <https://www.bmi.bund.de/SharedDocs/downloads/DE/publikationen/themen/moderne-verwaltung/open-data-strategie-der-bundesregierung.html>.
- [11] Rechtsrahmen zu Open-Data[EB/OL]. [2022-10-24]. <https://www.bmi.bund.de/DE/themen/moderne-verwaltung/open-government/open-data/open-data-node.htm>.
- [12] 2020/21 ODIN Annual Report[EB/OL]. [2022-11-12]. <https://odin.opendatawatch.com/Report/annualReport2020>.
- [13] Open Data in Europe 2020[EB/OL]. [2022-10-17]. <https://data.europa.eu/en/dashboard/2020>.
- [14] 司莉, 赵洁. 美国开放政府数据元数据标准及启示[J]. 图书情报工作, 2018, 62(3): 86-93.
- [15] 赵蓉英, 梁志森, 段培培. 英国政府数据开放共享的元数据标准——对Data.gov.uk的调研与启示[J]. 图书情报工作, 2016, 60(19): 31-39.
- [16] 黄如花, 李楠. 澳大利亚开放政府数据的元数据标准——对Data.gov.au的调研与启示[J]. 图书馆杂志, 2017(5): 87-97.
- [17] 翟军, 陶晨阳, 龙莎, 等. 欧盟开放数据的元数据标准DCAT-AP及启示[J]. 情报科学, 2019, 37(2): 102-110.
- [18] The German Open Data Portal uses the Interoperability Test Bed for data validation[EB/OL]. [2022-10-17]. <https://joinup.ec.europa.eu/collection/interoperability-test-bed-repository/solution/interoperability-test-bed/news/test-bed-support-govdatade>.
- [19] Gesetz zur Förderung der elektronischen Verwaltung[EB/OL]. [2022-11-12]. <http://www.gesetze-im-internet.de/egovg/>.
- [20] IT-Planungsrat. Standardisierungsbedarf “Metadatenstruktur für offene Verwaltungsdaten”[EB/OL]. [2022-10-30]. <https://www.it-planungsrat.de/beschluss/beschluss-2018-30>.
- [21] Vokabulare und Dokumente für DCAT-AP.de[EB/OL]. [2022-10-15]. <https://www.dcat-ap.de/def/>.
- [22] LÄMMEL P, DITTWALD B, BRUNS L, et al. Metadata Harvesting and Quality Assurance within Open Urban Platforms[J]. Journal of Data and Information Quality, 2020, 12(4): 1-20.
- [23] Open Data-Portal Bonn auf DCAT-AP.de umgestellt[EB/OL]. [2022-12-09]. <https://blog.bonn.de/digitaleverwaltung/?s=DCAT-AP.de>.
- [24] RDF DCAT to CKAN Dataset Mapping[EB/OL]. [2022-10-21]. <https://extensions.ckan.org/extension/dcat/>.
- [25] Redesign and new features for data.europa.eu[EB/OL]. [2022-11-17]. <https://data.europa.eu/en/news-events/news/redesign-and-new-features-dataeuropa.eu>.
- [26] 朱峥. 政府数据开放目录制度的问题及其应对之策[J]. 情报杂志, 2021, 40(12): 143-149.
- [27] 赵盼云, 耿骞, 柏林. 政府数据开放注册管理元数据研究[J]. 数字图书馆论坛, 2018(5): 52-61.
- [28] 国务院关于加强数字政府建设的指导意见[EB/OL]. [2022-11-17]. http://www.gov.cn/zhengce/content/2022-06/23/content_5697299.htm.
- [29] 张晓娟, 谭婧. 我国省级政府数据开放平台元数据质量评估研究[J]. 电子政务, 2019(3): 58-71.
- [30] Data Strategy of the Federal German Government[EB/OL]. [2022-10-28]. <https://www.bundesregierung.de/breg-de/suche/data-strategy-of-the-federal-german-government-1950612>.
- [31] 国务院办公厅关于印发全国一体化政务大数据体系建设指南的通知[EB/OL]. [2022-11-17]. http://www.gov.cn/zhengce/content/2022-10/28/content_5722322.htm.

- [32] Discovering millions of datasets on the web [EB/OL]. [2022-10-26]. <https://www.blog.google/products/search/discovering-millions-datasets-web/>.
- [33] Open Data Institute. Covid-19: Building an open and trustworthy data ecosystem [EB/OL]. [2022-10-30]. <https://theodi.org/project/building-an-open-and-trustworthy-data-ecosystem-for-covid-19>.
- [34] JAKUB K. DCAT-AP representation of czech national open data catalog and its impact [J]. Journal of Web Semantics, 2019 (55): 69-85.
- [35] KIRSTEIN F, KYRIAKOS S, BENJAMIN D, et al. Piveau: A Large-Scale Open Data Management Platform Based on Semantic Web Technologies [C] //European Semantic Web Conference 2020: The Semantic Web. Lecture Notes in Computer Science, 2020, 12123: 648-664.
- [36] KREMEN P, MARTIN N. Improving discoverability of open government data with rich metadata descriptions using semantic government vocabulary [J]. Journal of Web Semantics, 2019 (55): 1-20.

作者简介

翟军, 男, 1969年生, 博士, 教授, 研究方向: 开放数据、信息管理, E-mail: zhajun@dlmu.edu.cn。

毕智航, 男, 1998年生, 硕士研究生, 研究方向: 知识图谱。

董可新, 男, 1997年生, 硕士研究生, 研究方向: 开放数据、知识图谱。

沈立新, 女, 1970年生, 博士, 教授, 研究方向: 区块链、信息管理。

Metadata Standards and Application of Open Government Data in Germany

ZHAI Jun BI ZhiHang DONG KeXin SHEN LiXin

(School of Maritime Economics and Management of Dalian Maritime University, Dalian 116026, P. R. China)

Abstract: Germany is committed to becoming a leading open data country in the world, promoting innovation and development through open data. In particular, in 2021, Germany intensively introduced open data strategies, action plans and laws, which promoted metadata standard construction to show new characteristics and development trends. By means of web survey and literature analysis, this paper analyzes and introduces the DCAT-AP.de metadata standards for Open Government Data in Germany, and refines the experiences and practices for reference. For China, we should learn from the advanced experience in the policy and legal support of metadata standards, data governance and organizational guarantee, and the organic combination of localization and internationalization.

Keywords: Open Government Data; Metadata Standards; Data Catalogue; Data Governance; DCAT

(收稿日期: 2022-11-13)