

英国政府数据共享与开放的元数据标准建设及启示

翟 军 翟 玮 裴心童 李剑锋

【摘 要】[目的/意义]为促进政府数据的开放共享和高效使用,英国率先在元数据领域采纳开放标准,对其系统介绍可为我国政府数据资源的元数据标准建设提供一定的参考和借鉴。[方法/过程]采用文献分析和网络调查方法,分析和介绍2020年4月成立的英国政府数据标准局的第一批成果——共享和开放政府数据的元数据标准、实施指南和应用情况。[结果/结论]采纳开放标准代表着国际上元数据标准建设的最新发展方向。对我国而言,应在将元数据标准看作重要的数据基础设施、重视引进和鼓励使用开放标准、持续提升元数据管理的成熟度水平上借鉴先进经验。

【关键词】政府数据;元数据;开放标准;数据标准;数据共享;文献分析法;英国

【作者简介】翟军(ORCID:0000-0002-0525-2843)(1969-),男,大连海事大学航运经济与管理学院博士,教授,博士生导师,研究方向:开放数据和信息管理;翟玮(ORCID:0000-0002-0233-8131)(1996-),女,大连海事大学航运经济与管理学院硕士研究生,研究方向:开放数据;裴心童(ORCID:0000-0001-7864-439X)(1996-),女,大连海事大学航运经济与管理学院硕士研究生,研究方向:开放数据;李剑锋(ORCID:0000-0003-2548-6513)(1977-),男,大连海事大学航运经济与管理学院教授,研究方向:信息经济学(大连 116026)。

【原文出处】《情报杂志》(西安),2021.4.132~138,186

【基金项目】教育部人文社会科学研究规划基金项目“国家大数据战略下的政府开放数据的目录体系构建研究”(编号:17YJAZH115)和教育部2019年第一批产学研合作协同育人项目“信息管理专业大数据课程建设与改革”(编号:201901283023)。

0 引言

英国拥有欧洲最大的数据市场(Data Market),一直处于世界范围数据创新(Data Innovation)的前沿^[1]。据分析,2015-2020年大数据产业为英国带来约2 410亿英镑的收益,占GDP的2%左右,“数字经济”(Data Economy)已逐渐成为英国经济增长的重要引擎和数字经济的主要组成部分^[2]。

在发展数字经济的过程中,英国非常重视政府数据的作用,积极推动政府数据的共享、开放与应用^[3]。2017年3月,英国政府出台《数字化战略》(Digital Strategy),七大战略任务之一是发展数字经济、释放数据的价值^[4-5],包括更好地管理、共享和开放政府数

据,促进创造新的数据产品和服务。英国的《政府转型战略(2017-2020)》更是明确,通过开放政府数据、设立首席数据官和建立数据安全体系等措施保障政府数据的深度开发和利用^[6-7]。

2020年4月,英国政府成立“数据标准局”(Data Standards Authority, DSA),隶属于政府数字服务局(Government Digital Service, GDS),致力于数据标准的建设,以提升跨部门的数据共享和利用水平及保障数据质量^[8]。2020年8月,DSA公布第一批成果——共享和开放数据的元数据系列标准及实施指南,具有开放性和系统性的鲜明特征,代表着国际上元数据标准建设的发展方向^[9]。本文通过对该系列标准

的系统介绍和分析,提炼出有益的做法和经验,为我国各级政府数据资源的元数据标准建设提供借鉴和参考。

1 英国元数据标准建设情况

数据和元数据的“标准化”(Standardization)是各国推动政府数据开放共享的普遍做法和最佳实践^[9-10]。在元数据的标准和规范上,代表性的成果有美国的 DCAT-US(即 POD v1.1)、欧盟的 DCAT-AP、德国的 DCAT-AP、DE、澳大利亚的 Data. Gov. AU 元数据和英国的 CKAN 元数据等^[11-14]及面向地理领域的英国 GEMINI 元数据标准等^[14-15]。

英国连续五年(2013-2017年)居于万维网基金会的“开放数据晴雨表”(Open Data Barometer, ODB)全球评估的榜首,元数据标准的广泛应用对提升开放数据水平发挥了关键作用^[16]。欧盟数据门户网站(European Data Portal)对各国数据目录元数据质量的监测显示,英国国家数据目录在元数据互操作和促进数据再利用等质量维度上的表现较为突出^[17]。这次,英国又率先在政府数据共享和开放领域迈出了全面采纳“开放标准”(Open Standards)的步伐。

1.1 采纳开放标准 2017年底,在“创新英国”(Innovate UK,即英国政府技术战略委员会)的资助下,英国开放数据研究所(Open Data Institute, ODI)联合 W3C(World Wide Web Consortium, 万维网联盟)启动“数据的开放标准”(Open Standards for Data, OSD)项目(周期2017-2020年),旨在为政府和机构开发、采纳和实施开放标准提供指南和帮助^[18]。

ODI将“数据标准”定义为“关于数据的表示、格式、结构、传输、操作、使用和管理等的文档化协议”,而“开放的数据标准”则是以开放、协作和广泛共识的方式开发、批准和维护,任何人都可以免费或低成本访问、使用或共享的标准,能够帮助机构发布、访问、共享和使用质量更好的数据,更能为企业参与政府服务创造公平、高效的竞争环境。ODI认为,数据

的开放标准能够有效提高数据的互操作性、可比性、聚合性和连接性,与开放数据运动一起推动数据从封闭、共享和开放发展到广泛应用^[19]。ODI建议,数据治理和管理实践中应尽可能使用已有的开放标准,而不是开发新的标准。目前,OSD项目的主要成果包括开放标准的开发方法、选择指南和标准目录等。

在 OSD 项目成果的基础上,英国内阁办公室领导的“开放标准委员会”(Open Standards Board)制定了“开放标准的选择原则和流程”^[20],在开放数据领域已采纳和使用了“开放合同数据标准”和“国际援助透明度倡议数据标准”。

在开放标准委员会的指导下,数据标准局(DSA)确立三项开放标准作为英国政府数据共享和开放的元数据标准(见表1)^[21],分别是都柏林核心元数据(Dublin Core Schema,简称DC)、Schema.org 数据集词汇表和 W3C CSVW(CSV on the Web)标准。它们主要应用于通用的表格数据(Tabular Data),而在管理地理数据(包括 GeoJSON、KML 等格式)时,仍然使用原有的 GEMINI 标准。

1.2 政府数据共享的元数据标准 为实现高质量的政府数字服务,需要政府机构间的数据共享。DSA 选择都柏林核心元数据(DC)描述政府数据的主要原因是:①英国政府网站 GOV.UK 是数据共享的主要平台,而 DC 利于 Web 环境中数据的编目、校验和复用;②DC 是其他复杂元数据标准的基础,包括支持开放数据的元数据标准 DCAT(Data Catalog Vocabulary)和 Schema.org 数据集词汇表,有利于从数据共享到数据开放的平滑转换;③DC 元数据模式定义和元数据记录都支持机器可读的格式,如 XML Schemas、RDF Schemas、RDF/XML、JSON 和 RDFa 等,符合 W3C 的最佳实践^[10]。

在描述共享的政府数据时,DSA 要求提供:①发现元数据,使用的元素有 title(名称)、description(简介)

表1 英国政府数据共享和开放的元数据标准

序号	标准名称	应用领域	发布机构	发布时间
1	都柏林核心元数据	政府数据共享	Dublincore.org	XML/RDF 模式定义:1997-2020
2	Schema.org 数据集词汇表	政府数据开放	Schema.org	2012(V0.91)-2020(V9.0)
3	W3CCSVW 标准	描述表格数据	W3C(w3.org)	2015V1.0

和 identifier(标识符);②校验元数据,使用的元素有 expires(失效日期)、supersededBy(替换者,即 dct:isReplacedBy)、temporalCoverage(时间覆盖范围,即 dct:temporal)和 conformsTo(遵循的标准);③溯源元数据,使用的元素有 creator(创建者)、contributor(贡献者)和 dateCreated(创建日期);④使用元数据,使用的元素有 encodingFormat(编码格式)和 license(许可)等。

1.3 政府数据开放的元数据标准 为提高数据的 Web 可发现性,DSA 选择 Schema.org 数据集词汇表描述开放的数据集。

2011年6月,主要的搜索引擎公司 Google、Bing、Yahoo! 与 Yandex 联合发起 Schema.org 行动计划,共同开发和使用单一的“词汇表”(Vocabulary,即本体),为 Web 的结构化数据标记提供统一的模式定义^[21]。从 2012 年至今,Schema.org 词汇表的版本从 V0.91 升级到 V9.0,从最初的 297 个类、187 个属性,发展到 626 个类、901 个属性。从 2013 年 4 月的 V1.0a 开始,Schema.org 联合 W3C 进行 Web Schemas/Datasets 部分的扩展,到 2019 年 4 月的 V3.5 版本形成了稳定的数

据集词汇表,包括 DataCatalog(数据目录)、Dataset(数据集)、DataDownload(可下载的数据/文件)和 WebAPI(数据接口/服务)四个核心类,见表 2。

英国政府数字服务局(GDS)历来重视使用 Schema.org 词汇表标注政府网站的内容,以提高搜索引擎的搜索效果^[23]。例如,GOV.UK 已使用 Schema.org 的 Special Announcement 词汇表描述新冠疫情信息,使得在谷歌(Google)搜索中突出显示 Covid-19 的相关内容。对于数据集的描述,GDS 和 DSA 建议的元数据元素(即词汇表中类的属性)见表 3,这也是谷歌“数据集搜索”(Dataset Search)引擎工具所要求的^[24]。

1.4 Tabular 数据的元数据标准 英国开放标准委员会将 CSV(Comma-Separated Values,逗号分隔值)文件格式确立为政府 Tabular 数据(即表格数据)的首选格式。进一步,DSA 将 W3C CSVW 标准确立为描述 Tabular 数据内容和结构的元数据标准,以帮助用户更好地理解和使用数据。

W3C 于 2013 年成立 CSV 工作组(CSV on the Web Working Group),联合英国开放数据研究所(ODI)于

表 2 Schema.org 数据集词汇表的核心类

序号	类的名称	类的 URL	对应的 DCAT 类	中文名称
1	DataCatalog	https://schema.org/DataCatalog	dcat: Catalog	数据目录
2	Dataset	https://schema.org/Dataset	dcat: Dataset	数据集
3	DataDownload	https://schema.org/DataDownload	dcat: Distribution	数据文件
4	WebAPI	https://schema.org/WebAPI	dcat: DataService	数据接口

表 3 Schema.org 数据集词汇表的主要属性

序号	属性	定义域	值域	对应的 DC/DCAT 属性	中文名称
1	name	Dataset	Text	dct: title	名称
2	description	Dataset	Text	dct: description	摘要/简介
3	keywords	Dataset	Text	dcat: keyword	关键字
4	about	Dataset	Thing	dcat: ththeme	主题/领域
5	spatialCoverage	Dataset	Text 或 Place 类	dct: spatial	空间覆盖范围
6	url	Dataset	URL	dcat: landingPage	在线网址
7	datePublished	Dataset	DateTime	dct: issued	发布日期
8	dateModified	Dataset	DateTime	dct: modified	更新日期
9	publisher	Dataset	Organization 类	dct: publisher	数据发布者
10	distribution	Dataset	DataDownload	dcat: distribution	数据下载
11	includedInDataCatalog	Dataset	DataCatalog	无	所在目录
12	license	Dataset	Text 或 URL	dct: license	开放许可
13	contentUrl	DataDownload	URL	dcat: downloadURL	下载地址

2015年12月-2016年2月发布了Web中CSV格式数据的系列标准^[25],见表4。

为使得用户的应用程序能够通过读取和分析元数据,自动实现CSV数据的校验、转换、聚合、显示和搜索功能,CSVW标准明确定义了CSV数据的结构和数据类型,并建立描述这种结构的“词汇表”(即元数据模型),含有的词汇诸如csvw:TableGroup、csvw:Table、csvw:Row、csvw:Column、csvw:Cell、csvw:Schema、csvw:Datatype、csvw:format等,形成W3C最佳实践所推荐的“结构型元数据”(Structural Metadata)^[10]。在描述CSV对象(如csvw:TableGroup、csvw:Table)的一般属性(如title、keyword等)时,仍使用DC、DCAT或Schema.org元数据。

2 元数据标准的实施指南

伴随着正式采纳“开放标准”及配合《政府转型战略(2017-2020)》的“高效使用数据”的战略目标,英国政府数字服务局(GDS)和数据标准局(DSA)发布了

上述元数据标准的实施指南(见表5的前3个),同时更新了一些指南(见表5的后4个)配合这些指南的使用。

指南要求政府机构发布数据的同时要提供完备的元数据。除了面向人的阅读,还应提供机器可读的元数据:元数据可以与数据在同一个文件中,也可以放到单独的文件中,或者嵌入到数据集所在的网页。当数据与元数据不在一起时,要设置从数据到元数据的链接。例如,政府网站GOV.UK和开放数据网站DATA.GOV.UK是数据共享和开放的主要渠道,其中每个数据集的机器可读的元数据记录采用了JSON-LD格式,以脚本的方式嵌入到网页的HTML代码中;数据目录网站ckan.publishing.service.gov.uk/dataset则为每个数据集提供了四种格式的元数据文件,分别是XML、JSON-LD、TTL和N3。

指南对数据集的标识符也给出了明确的规定。在数据共享时,标识符可以是局部的,如“362857580”,而在数据开放时,标识符应是全局的,建议采用URL

表4

W3C CSVW 系列标准

序号	标准名称	说明
1	Web中Tabular数据的模型与元数据	描述Tabular数据和Csv格式的含义,定义Tabular数据模型和元数据类型,及如何表示和利用元数据
2	Tabular数据的元数据词汇表	明确元数据文档的结构和格式,定义JSON-LD格式的元数据描述表组、表、列等对象时需要的属性(即词汇),及各种数据类型等
3	Csv格式到JSON格式的转换	将Csv格式数据转换为JSON数据的方法和处理流程
4	Csv格式到RDF格式的转换	将Csv格式数据转换为RDF数据的方法和处理流程
5	Web中的Csv数据:案例和需求	通过分析统计数据、科学数据、出版数据和支出数据等25个案例,提出Csv的数据模型需求、解析需求 and 应用需求
6	在HTML网页中嵌入Tabular元数据	在网页中提供元数据的方法,包括链接到外部文件和嵌入JSON-LD脚本
7	web中的Csv:入门指南	通过示例介绍如何为Csv数据提供各种粒度的元数据:表格、行/列和单元粒度等

来源: https://www.w3.org/2013/csvw/wiki/Main_Page。

表5

元数据标准的实施指南

序号	指南名称	发布/更新日期	说明
1	共享数据集的元数据指南	2020年8月	如何使用DC描述共享的表格形式的政府数据集(地理数据除外),包括元数据的存储方式、必需的元数据项等
2	开放数据集的元数据指南	2020年8月	如何使用Schema.org数据集词汇表描述开放的数据集,以提高数据集的搜索引擎可发现性
3	发布表格数据指南	2020年8月	共享和开放数据时,如何为Csv数据文件提供W3CCSVW元数据
4	政府数据和技术的开放标准	2020年8月	开放标准原则,政府数据交换和发布的开放标准列表
5	表格数据标准指南	2020年8月	发布表格形式的政府数据时,首选Csv格式,应遵循RFC4180标准
6	日期数据格式指南	2020年7月	遵循ISO 8601:2004标准,以机器可读的格式表达日期数据,格式为YYYY-MMDD-THH:MM-SS
7	持久化可解析标识符指南	2020年7月	使用HTTP1.1和URL标识开放的数据集

(Uniform Resource Locator, 统一资源定位器) 且是 HTTP 协议可解析的。目前, GOV.UK 和 DATA.GOV.UK 都为每个数据集分配了持久的、静态的 URL, 可直接定位到数据集所在的网页。

指南还特别强调, 在创建、发布和使用元数据时应遵守英国《数据保护法(2018)》(Data Protection Act 2018) 和《数据伦理框架》(Data Ethics Framework) 对数据安全、个人隐私保护和数据使用的伦理道德的规定。

3 元数据标准的应用

随着开放数据网站和数据集数量的快速增长, 谷歌(Google)于2018年9月推出专业搜索引擎“数据集搜索”(测试版), 致力于打造全新的国际数据生态系统。该搜索引擎是新一代的智能搜索引擎, 只有以 Schema.org 数据集词汇表标记的网页才能被采集到专门的索引库^[24]。2020年1月, 其正式版发布, 收录了全世界数千个网站中的约2 500万个数据集^[26]。

英国数据标准局(DSA)采纳 Schema.org 词汇表的主要目的就是提高数据的 Web 搜索引擎可发现性, 促进政府数据的大范围流动, 并已取得了良好的效果。例如, 当使用谷歌数据集搜索引擎工具搜索“新冠疫情”(COVID-19)的相关数据时, 可以找到100多个来自英国的数据集, 它们广泛分布在各级政府网站、国家医疗服务体系(National Health Service, NHS)网站和各类开放数据网站中, 网页标注的方式主要是 JSON-LD 和微数据^[27], 实例见表6。

图1显示了布里斯托市开放数据网站中一个数

```
<script type="application/ld+json">
{
  "@context": "http://schema.org/",
  "@type": "Dataset",
  "name": "Covid-19 cases",
  "description": "This is a subset of the lab-confirmed cases of COVID-19 data published by PHE on the data.gov.uk website. ....",
  "url": "https://opendata.bristol.gov.uk/explore/dataset/covid-19-cases/",
  "dateModified": "2020-09-09T17:00:46.464768+00:00",
  "keywords": ["Coronavirus", "COVID", "COVID-19"],
  "license": "http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/",
  "distribution": [
    {
      "@type": "DataDownload",
      "encodingFormat": "CSV",
      "contentUrl": "https://opendata.bristol.gov.uk/explore/dataset/covid-19-cases/download?format=csv",
      "format": "csv"
    }
  ]
}
```

图1 布里斯托市开放数据网站数据集 Schema.org 标注实例
来源: <https://opendata.bristol.gov.uk/explore/dataset/covid-19-cases>。

据集的元数据实例, 标注代码的格式是 JSON-LD, 使用的 Schema.org 元数据项有 name、description、url、keywords 等(见表3)。从元数据记录可见这个数据集含有 CSV 格式的数据文件, 数据集所在的网页也给出了 CSV 数据的结构元数据, 定义了表格的 area_name、specimen_date 和 geo_shape 等字段及其数据类型。

面向数据使用者和普通用户, 数据集搜索引擎工具大幅度提高了数据搜索的便捷性和准确性。调查显示, 截至2019年下半年, 42%的欧洲数据门户网站的访问者借助了搜索引擎的引导^[28]。在开放数据领域, 采纳和使用 Schema.org 数据集词汇表等开放标准, 能够帮助英国加速形成国际化的数据生态系

表6 英国发布的新冠疫情开放数据集实例

序号	数据集名称	所在网站	标注方式
1	COVID-19: number of outbreaks in care homes	英国政府网站 www.gov.uk	JSON-LD
2	Coronavirus(COVID-19)Deaths	国家开放数据网站 data.gov.uk	JSON-LD
3	Travelpac: travel to and from the UK	国家统计局网站 www.ons.gov.uk	JSON-LD
4	Coronavirus Shielded Patient List Summary Totals	国家医疗服务体系(NHS)数字服务中心 digital.nhs.uk	微数据
5	Covid-19 cases	布里斯托市开放数据网站 opendata.bristol.gov.uk	JSON-LD
6	Coronavirus(COVID-19)In Camden And Surrounding Boroughs	卡姆登市开放数据网站 opendata.camden.gov.uk	JSON-LD

来源: <https://datasetsearch.research.google.com/search?query=UK COVID-19>。

统(Data Ecosystem, 见图2),助力实现《国家数据战略》(National Data Strategy, NDS)确立的“数据的国际流动”的目标^[29],在数据驱动的抗击新冠疫情的国际合作上更是发挥着积极作用。

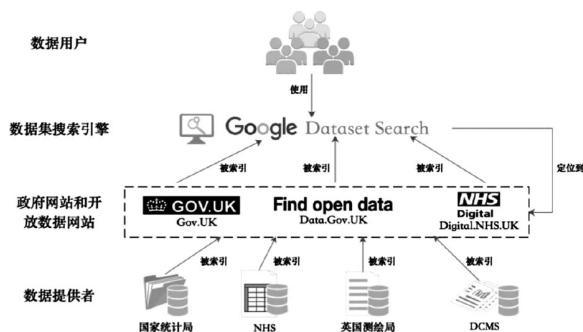


图2 英国开放数据的生态系统

4 对我国的启示

在国家大数据战略的引领下,我国各级政府正在稳步推进数据开放工作。根据复旦大学《中国地方政府数据开放报告》,截至2020年4月底,我国已有130个省级、副省级和地级政府上线了开放数据平台^[30]。近两年,上海、浙江、哈尔滨和青岛等省市政府先后出台《公共数据开放管理办法》,确立了“有序开放、安全可控、统一标准”等工作原则^[31]。

在元数据标准的建设上,山东、广东和贵州等率先研制和发布了开放数据元数据的地方标准^[30],如山东省的《政务信息资源目录第2部分:核心元数据》(DB37/T3521.2-2019)、广东省的《电子政务数据资源开放数据技术规范》(DB44/T2110-2018)和贵州省的《政府数据开放数据核心元数据》(DB52/T1407-2019)。2020年4月,国家标准化管理委员会发布了《信息技术大数据政务数据开放共享第1部分:总则》(GB/T38664.1-2020)、《第2部分:基本要求》(GB/T38664.2-2020)和《第3部分:开放程度评价》(GB/T38664.3-2020)三项政府数据开放共享领域的国家标准,于2020年11月1日起实施^[32]。

但我国还缺乏国家层面的开放政府数据的统一元数据标准,元数据标准建设滞后于实践需求,也落后于国际先进水平。相应地,我国公共数据开放共享总体上还处于发展阶段,在ODB上的排名仅为第24位^[16],实践中还需要积极借鉴国际上的先进经验^[33-35]。在此背景下,应重点从以下三方面学习和借

鉴英国的做法和经验。

4.1 将元数据标准建设提升到数据基础设施的高度 在大数据和开放数据时代,英国政府将“数据基础设施”(Data Infrastructure)的重要程度提升到与传统的基础设施(如交通、通信等)相同的高度^[36]。ODI认为,数据基础设施不仅包括数据资产本身,还包括如何管理和使用数据的标准和指南及管理维护这些资产的组织、人员、设施和技术等^[19]。在政府数字服务局(GDS)的领导下,英国政府数据标准局(DSA)将在未来3年建立一个以数据标准为核心的“数据基础设施”,以帮助政府部门提供以用户为中心的、跨越组织机构边界的数字服务^[37],而政府数据开放共享的元数据标准和指南则是这个数据基础设施的第一个“构造块”。

我国正在推动各类新型基础设施的建设。在“新基建”中,数据中心(如科学数据中心、政府数据中心等)等数据基础设施是投资的重点之一,连续3年保持着30%以上的增长^[38]。在数据中心的建设中,应重视数据标准和元数据标准的同步开发和应用,打破“数据孤岛”,实现数据资源之间的连接和互操作,为“新基建”在各个领域的广泛应用提供海量数据的支撑^[38]。

4.2 重视引进开放标准和融入国际数据生态系统

英国历来重视在政府数字服务中使用国际标准和开放标准。2012年6月出台的《开放数据白皮书》将“使用开放标准”确立为开放数据的十四项原则之一^[36]。2020年9月9日,英国数字、文化、媒体和体育部(DCMS)颁布《国家数据战略》(National Data Strategy, NDS),将“在安全的前提下支持数据的国际流动”确立为5个战略目标之一^[29],保障的措施包括与各国合作开发“共享标准”,这次将DC和Schema.org数据词汇表等开放标准采纳为国家层面的元数据标准只是迈出的第一步。

在应对新冠疫情的国际合作中,通过开放数据和开放标准构建开放可信的国际数据生态系统日益得到各国的重视^[39]。我国的开放政府数据,特别是新冠疫情数据,也是国际数据生态系统的重要一环。但还存在着一些问题影响着我国在国际合作上的主动权和话语权,如数据标准缺失^[40]、元数据与国际标准(Schema.org和DCAT)不兼容和通过搜索引擎

(如百度、微软 Bing 和谷歌等)难以发现和定位到开放数据集等。

因此,我国在开展大数据和开放数据领域新标准研制工作的同时,应重视引进和鼓励使用国际上成熟的开放标准,同时积极参与数据标准开发和推广使用的国际合作,如 W3C 的 Web 数据标准化工作已经吸引了英国、德国、加拿大和澳大利亚等国家的几十家合作机构^[41]。

4.3 加强元数据治理、提高元数据管理的成熟度水平 从2011年至今,英国政府已连续实施四轮“开放政府国家行动计划”(Open Government National Action Plan)^[42],数据标准和开放标准建设是贯穿其中的重点内容之一。这次采纳元数据领域的开放标准是英国政府持续开展元数据治理、不断提高元数据管理水平的标志性成果之一。

政府元数据管理的成熟度水平从低到高分5个级别^[43],分别是:①第一级别,没有认识到元数据的重要性,随意使用元数据、数据模型和词汇表等,缺乏文档化的元数据,导致互操作问题;②第二级别,已实现元数据的文档化,但还不是集中管理和结构化的;③第三级别,元数据经过充分的文档化、可被重用,但还不是机器可读的;④第四级别,使用开放且可重用的元数据,元数据集中管理,提供机器可读的文档或 API;⑤第五级别,使用语义 Web 技术(如词汇表/本体、RDF 和关联数据等)定义和描述元数据,元数据是关联且开放的。目前,英国、美国、澳大利亚和欧盟的开放数据元数据管理基本处于第四或第五级别^[11-14],其中英国的成熟度水平无疑是最高。

我国各地方政府的元数据还存在着互操作水平低^[12]、不是机器可读的和数据质量不高^[44]等问题,是政府数据治理的薄弱环节。研究表明,“治理能力”是开放政府数据的关键制约因素^[45],我国应构建跨部门的政府数据治理框架^[46]。因此,政府部门应通过有效的元数据治理,持续提高管理的成熟度水平和元数据质量,为开放政府数据的可持续健康发展打下坚实的基础。

5 结语

英国是世界范围开放政府数据运动的领跑者,其数据战略、政策法规、行动计划、平台和标准建设

等一直是各国的典范。元数据标准是保障政府数据开放共享的关键基础设施,得到了各国的普遍重视。近年来,英国相继成立开放标准委员会和数据标准局,加大了元数据领域开放标准的采纳、建设和推广力度,为数字经济时代实施新的国家数据战略和构建国际化的数据生态系统提供了有力保障。本文在对该系列标准进行系统介绍和分析的基础上,提炼出了对我国有益的启示。

后续研究将继续关注英国新的数据战略框架下的数据标准建设、数据保护法规、数据伦理框架和数据驱动的人工智能等内容,为我国数字经济和开放数据的政策、法规和标准建设提供更多的建议和参考。

参考文献:

- [1]Department for Digital, Culture, Media & Sport. Data-unlocking the power of data in the UK economy and improving public confidence in its use[EB/OL].[2020-08-31]. <https://www.gov.uk/government/publications/uk-digital-strategy/7-data-unlocking-the-power-of-data-in-the-uk-economy-and-improving-public-confidence-in-its-use>.
- [2]CEBR & SAS. The value of big data and the internet of things to the UK economy[EB/OL].[2020-08-31]. https://www.sas.com/content/dam/SAS/en_gb/doc/analystreport/cebr-value-of-big-data.pdf.
- [3]中国信息通信研究院. 全球数字经济新图景(2019年)[EB/OL].[2020-09-01]. http://www.caict.ac.cn/kxyj/qwfb/bps/201910/t20191011_214714.htm.
- [4]Department for digital, culture, media & sport. UK digital strategy[EB/OL].[2020-09-01]. <https://www.gov.uk/government/publications/uk-digital-strategy>.
- [5]陈美. 政府开放数据的隐私风险评估与防控:英国的经验[J]. 中国行政管理, 2020(5): 153-159.
- [6]Cabinet Office, Government Digital Service. Government transformation strategy 2017 to 2020[EB/OL].[2020-09-01]. <https://www.gov.uk/government/publications/government-transformation-strategy-2017-to-2020>.
- [7]张晓, 鲍静. 数字政府即平台:英国政府数字化转型战略研究及其启示[J]. 中国行政管理, 2018(3): 27-32.
- [8]Data standards authority(DSA) [EB/OL].[2020-09-01]. <https://www.gov.uk/government/groups/data-standards-authority>.
- [9]European Data Portal(EDP). The data standards authority (DSA) has published a collection of metadata standards for the sharing and publication of data[EB/OL].[2020-09-01]. <https://www.europeandataportal.eu/en/news/uk-governments-data-standards-authority-publishes-metadata-standards>.
- [10]W3C. Data on the Web Best Practices[EB/OL].[2020-09-01]. <https://www.w3.org/TR/dwbp/>.

[11]于梦月,翟军,林岩,等.美国政府开放数据的元数据标准及其启示:目录聚合的视角[J].情报杂志,2017(12):145-151.

[12]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.

[13]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.

[14]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.

[15]夏姚璜.开放政府数据平台的地理空间元数据标准研究[J].图书馆建设,2018(8):40-46,53.

[16]World Wide Web Foundation. Open data barometer[EB/OL].[2020-12-08]. <https://opendatabarometer.org/leadersedition/report/>.

[17]European Data Portal(EDP). Metadata quality assurance[EB/OL].[2020-12-08]. <https://www.europeandataportal.eu/mqa>.

[18]Open Data Institute. Documenting the development of open standards for data[EB/OL].[2020-09-02]. <https://theodi.org/article/documenting-the-development-of-open-standards-for-data/>.

[19]Open Data Institute. When to use open standards for data[EB/OL].[2020-09-02]. <https://standards.theodi.org/introduction/when-to-use-open-standards-for-data/>.

[20]Cabinet Office. Open standards principles[EB/OL].[2020-09-03]. <https://www.gov.uk/government/publications/open-standards-principles/open-standards-principles>.

[21]Data Standards Authority. Metadata standards for sharing and publishing data[EB/OL].[2020-09-03]. <https://www.gov.uk/government/collections/metadata-standards-for-sharing-and-publishing-data>.

[22]Guha R V, Brickley D, Macbeth S. Schema. Evolution of structured data on the Web[J].Communications of the ACM, 2016, 59(2): 44-51.

[23]Government Digital Service. What data we expose as schema. org structured data[EB/OL].[2020-09-06]. <https://docs.publishing.service.gov.uk/manual/schemas.html#dataset>.

[24]Brickley D, Burgess M, Noy N. Google dataset search: Building a search engine for datasets in an open Web ecosystem [C]. The World Wide Web Conference, ACM, 2019: 1365-1375.

[25]CSV on the Web working group[EB/OL].[2020-09-08]. https://www.w3.org/2013/csvw/wiki/Main_Page.

[26]Discovering millions of datasets on the Web[EB/OL].[2020-09-10]. <https://www.blog.google/products/search/discovering-millions-datasets-web/>.

[27]贾君枝,王醒.基于微数据的语义标注应用研究[J].情报理论与实践,2016,39(2):54-58.

[28]European Data Portal(EDP). Sustainability of(Open) data portal infrastructures: A distributed version control approach to creating portals for reuse[EB/OL].[2020-09-14]. https://www.europeandataportal.eu/sites/default/files/sustainability-data-portal-infrastructure_6_distributed-version-control.pdf.

[29]Department for Digital, Culture, Media & Sport. National data strategy[EB/OL].[2020-09-14]. <https://www.gov.uk/government/publications/uk-national-data-strategy/national-data-strategy>.

[30]复旦大学移动与数字治理实验室.中国地方政府数据开放报告(2020上半年)[EB/OL].[2020-09-15].<http://ifopendata.fudan.edu.cn/report>.

[31]上海市公共数据开放暂行办法[EB/OL].[2020-09-15]. <http://www.sh.gov.cn/nw2/nw2314/nw2319/nw12344/u26aw62638.html>.

[32]国家标准化管理委员会.信息技术大数据政务数据开放共享第1部分:总则[EB/OL].[2020-09-15].<http://std.samr.gov.cn/gb/search/gbDetailed?id=A47A713B767914ABE05397BE0A0ABB25>.

[33]孟显印,杨超.我国开放政府数据应用开发的现状与问题[J].情报杂志,2020,39(3):163-171,197.

[34]周文泓,夏俊英.加拿大政府开放数据的特点研究及启示[J].情报理论与实践,2018,41(4):150-154,138.

[35]杨晶,康琪,李哲.美国《联邦数据战略与2020年行动计划》的分析及启示[J].情报杂志,2020,39(9):150-156,94.

[36]翟军,翁丹玉,袁长峰,等.英国政府开放数据的“国家信息基础设施”建设及启示[J].情报科学,2017,35(6):107-114.

[37]Rosalie Marshall. Introducing the government data standards authority[EB/OL].[2020-09-15].<https://dataingovernment.blog.gov.uk/2020/08/27/introducing-the-government-data-standards-authority/>.

[38]吴忠泽.新基建 新科技 引领中国数据中心产业高质量发展[EB/OL].[2020-12-09]. <http://news.idcquan.com/news/182643.shtml>.

[39]Open Data Institute. Covid-19: Building an open and trustworthy data ecosystem[EB/OL].[2020-09-16].<https://theodi.org/project/building-an-open-and-trustworthy-data-ecosystem-for-covid-19>.

[40]周林兴,徐承来,宋大成.重大疫情灾害中政府数据开放模式研究——以新型冠状病毒肺炎疫情为实证分析[J].现代情报,2020,40(6):3-18.

[41]W3C. W3C study of practices and tooling for Web data standardization[EB/OL].[2020-09-16].<https://www.w3.org/2017/12/odi-study/>.

[42]Department for Digital, Culture, Media & Sport. Open government[EB/OL].[2020-09-16].<https://www.gov.uk/government/collections/open-government>.

[43]European Data Portal(EDP).Sustainability of(Open) data portal infrastructures: Open data portal assessment using user-Oriented metrics[EB/OL].[2020-09-16].https://www.europeandataportal.eu/sites/default/files/sustainability-data-portal-infrastructure_5_open-data-portal-assessment.pdf.

[44]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.

[45]吴金鹏,韩啸.开放政府数据何以成功?——生态关系理论视角的跨国实证研究[J].图书馆论坛,2020(8):87-95.

[46]左美云,王配配.数据共享视角下跨部门政府数据治理框架构建[J].图书情报工作,2020,64(2):116-123.