

国外政府门户开放重大公共卫生事件相关数据现状分析*

李赞梅 吴思竹 王安然

(中国医学科学院/北京协和医学院医学信息研究所 北京 100020)

〔摘要〕 以受新型冠状病毒肺炎疫情影响较大及政府数据开放水平较高的 5 个国家(美国、印度、英国、加拿大和澳大利亚)为研究对象,从数据发布时间、发布机构、数据主题、开放等级、使用许可几方面调查分析国外政府新冠肺炎疫情数据开放现状及主要问题。

〔关键词〕 政府数据开放;新型冠状病毒肺炎;现状分析

〔中图分类号〕 R-058 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2021.03.001

Analysis of the Current Situation of Data Related to Major Public Health Events Opened by Foreign Governments LI Zanmei, WU Sizhu, WANG Anran, Institute of Medical Information, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100020, China

〔Abstract〕 Five countries (the United States, India, the United Kingdom, Canada and Australia) that are most affected by COVID-19 epidemic and have the highest level of government data openness are selected as the study objects, and the current situation and major problems of COVID-19 epidemic data opening of foreign governments are investigated and analyzed from the aspects of data release time, release agency, data subject, opening level and usage license.

〔Keywords〕 government data opening; COVID-19; status analysis

〔修回日期〕 2020-12-22

〔作者简介〕 李赞梅, 硕士, 助理研究员, 发表论文 20 篇; 通讯作者: 吴思竹, 博士, 副研究员。

〔基金项目〕 中国医学科学院北京协和医学院医学信息研究所/图书馆新冠肺炎疫情专项“全球新冠肺炎疫情危机下重大传染病疫情相关政府开放数据资源目录研究”(项目编号: 2020XG1125); 国家社会科学基金青年项目“‘国家大数据战略’背景下健康医疗开放数据资源分类体系构建与应用研究”(项目编号: 17CTQ020)。

1 引言

新型冠状病毒肺炎(以下简称新冠肺炎)疫情持续蔓延,严重影响全球稳定、社会发展以及人类生存与生活。当前世界多国普遍进入疫情防控新常态,政府治理和应急管理能力受到新的挑战。政府开放新冠肺炎疫情相关数据,为公众提供及时、准确的信息,对于疫情防控和社会经济发展至关重要。我国较早遭受新冠肺炎疫情冲击,国家、各省市县政府及卫生部门均及时开放疫情相关数据^[1-3],向公众提供确诊病例统计、发热门诊机构清单、个案详情等。目前全球多地仍深受新冠肺炎疫情困

扰，部分国家疫情甚至出现反复^[4]。本文以新冠肺炎疫情相关政府开放数据为对象，选取受新冠肺炎疫情冲击较大和政府数据开放排名前列的国家为样本，从开放数据数量、发布时间、发布机构、主题分布、开放等级和开放许可几方面对国外政府疫情数据开放情况进行调查研究，分析存在的问题。

2 研究对象与方法

2.1 研究对象

世界卫生组织发布的新冠肺炎疫情统计数据^[5]显示截至 2020 年 7 月 21 日全球累计确诊病例超 1 400 万（其中死亡逾 61 万）。美国确诊人数超 374 万（占全球 26.7%），是疫情最严重的国家。累计确诊病例前 10 位的其他国家，除英国外（第 9 位），巴西、印度、俄罗斯、南非等 8 国均为发展中国家。政府数据开放评估方面，国内外已有大量研究与项目^[6]，代表性的如世界银行开放数据准备度^[7]，经济合作与发展组织（Organization for Economic Co-operation and Development, OECD）开放政府数据指数^[8]，互联网基金会开放数据晴雨表（Open Data Barometer, ODB）^[9]。其中开放数据晴雨表较具普遍性和影响力，在其最新公布的报告中，加拿大、英国、澳大利亚位居前 3 位，美国第 9 位，中国第 24 位。综合新冠肺炎疫情严重程度及政府数据开放水平两方面因素，选取美国、印度、英国、加拿大和澳大利亚 5 个国家纳入调研范围。各国疫情及政府数据开放基本情况，见表 1。

国家	疫情基本情况		政府数据开放基本情况	
	排名	确诊病例数	ODB 排名	国家平台网址
美国	1	3 748 248	9	Data. gov
印度	3	1 155 191	16	data. gov. in
英国	9	295 376	2	Data. gov. uk
加拿大	21	110 338	1	open. canada. ca
澳大利亚	72	12 069	3	data. gov. au

注：各国疫情数据统计时间为 2020 年 7 月 21 日。

以美、印、英、加、澳 5 国政府数据开放统一平台上发布的新冠肺炎疫情相关数据为对象，从数据发布时间、发布机构、主题、开放等级和开放许

可 5 个方面调研新冠肺炎疫情事件中政府数据开放发展实践情况。采取自动抓取与人工采集相结合的方式获取数据，采集时间为 2020 年 7 月 21 日。

2.2 研究方法

采用以定量分析为主的研究方法。其中数据主题采用《中文医学主题词表》（Chinese Medical Subject Headings, CMeSH）^[10]进行描述和分类。由两位具有多年生物医学数据标注经验的调查人员根据数据集元数据和/或内容信息，选择最核心的一个主题进行人工标注，不一致的经讨论形成一致结论；数据开放等级评估采用开放数据五星评估方案（5-star Open Data）^[11]。该方案从数据本身可访问性和机器可读性技术层面，由低到高将开放数据划分为 1~5 星共 5 个等级。1 星表示数据可在网络上获取，2 星表示数据已结构化，3 星表示实现以非专用格式提供，4 星表示数据采用统一资源标识符（Uniform Resource Identifier, URI）标记，5 星即开放关联数据。除加拿大政府开放数据已采用五星评估方案并可直接获取数据开放等级外，其余 4 个国家数据开放等级均由调查人员根据数据格式判断确定。

3 结果与分析

3.1 概述

截至 2020 年 7 月 21 日美、印、英、加、澳 5 个国家政府开放数据平台上共发布 243 个新冠肺炎疫情相关数据集。美国开放数据集数量最多，占 5 国总数的 48.1%（117 个）。加拿大次之，开放 61 个数据集，占 25.1%。印度和澳大利亚数据集数量相当，分别有 26 个和 25 个。英国最少，仅有 14 个。

3.2 数据发布时间

按月份统计的 5 国政府开放数据发布时间，见图 1。除印度外，美、英、加、澳 4 国均在 3 月开始发布疫情相关数据集。美国最早于 2020 年 3 月 12 日发布第 1 个数据集，为华盛顿州疫情动态数据^[12]。2020 年 4 月以后美国开放数据数量增长较快（2020 年 7 月数据统计截至 7 月 21 日），这与其

国内确诊病例快速增长的趋势较为一致。2020年4月加拿大集中发布一批来自国家统计局关于疫情对社会经济影响的数据,美国开放来自多个地方政府的疫情动态数据。印度于2020年5月发布首个数据集,6、7月数据数量持续增长,这与5月以后印度疫情迅速蔓延存在关联。澳大利亚疫情早期防控效果较好,各月开放的数据集数量均在较低水平。英国开放的数据数量各月也一直处于较低水平。

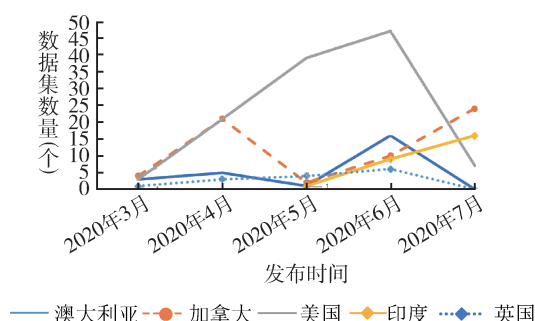


图1 各国新冠肺炎疫情相关数据集发布时间

3.3 数据发布机构

5 国政府开放新冠肺炎疫情相关数据集发布机构分布,见图2。美国数据集来源最为多样,除卫生部门外,有企业、国际贸易、标准研发、交通运输等机构提供的数据。联邦政府层面,美国疾病控制与预防中心、中小企业发展署(Small Business Administration, SBA)发布的数据集最多(29个和19个),也有来自联邦运输统计局(Bureau of Transportation Statistics, BTS),国家标准与技术研

究院(National Institute of Standards and Technology, NIST),国际贸易委员会(International Trade Commission, ITC)的数据集。美国过半的数据集(62个, 53%)来自10个州、市或郡级地方政府数据平台,如旧金山、康涅狄格、纽约和华盛顿。加拿大数据集绝大多数(85.2%, 52个)来自国家统计局,为关于疫情对社会经济影响的数据。其余数据全部来自公共卫生部(Public Health Agency of Canada)、全球事务部(Global Affairs)、就业及社会发展部(Employment and Social Development)以及移民、难民和公民部(Immigration, Refugees and Citizenship)4个部门。英国政府收集的数据集90%来自4个地方政府(12个),包括大伦敦政府(Greater London Authority)以及3个市议会(Leeds、York、Calderdale),中央政府发布的两个数据集来自食品标准局(Food Standards Agency)以及商业、能源和产业战略部(Department for Business, Energy and Industrial Strategy)。与英国类似,澳大利亚数据集除1个来自卫生部外,其余24个(96%)均来自新南威尔士州政府及卫生厅。印度数据88.5%(23个)来自负责全国卫生与人口工作的卫生和家

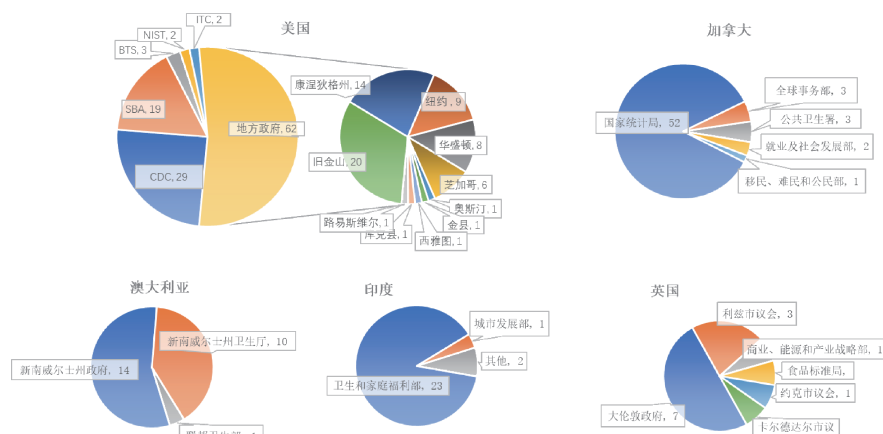


图2 各国新冠肺炎疫情相关数据集发布机构/来源分布

3.4 数据主题分布

5 国政府开放的新冠肺炎疫情相关数据集被归入 21 个主题，见表 2。

表 2 5 国政府新冠肺炎疫情相关数据集主题分布（个）

主题	澳大利亚	加拿大	美国	印度	英国	总计
人口统计学	11	11	54	26	6	108
经济学	-	37	4	-	1	42
政府补助	-	3	25	-	2	30
救济工作	14	-	2	-	1	17
卫生设施	-	1	7	-	-	8
运输	-	2	4	-	-	6
体育运动	-	-	5	-	-	5
临床实验室技术	-	-	3	-	-	3
卫生服务	-	-	2	-	1	3
卫生服务需求	-	3	-	-	-	3
图书馆服务	-	-	3	-	-	3
登记	-	3	-	-	-	3
危险因素	-	-	1	-	1	2
精神卫生	-	-	2	-	-	2
文献工作	-	-	2	-	-	2
管理审计	-	-	-	-	1	1
环境污染	-	-	1	-	-	1
计算机安全	-	1	-	-	-	1
卫生经费支出	-	-	1	-	-	1
隐私权	-	-	1	-	-	1
研究	-	-	-	-	1	1
总计	25	61	117	26	14	243

涉及病例统计、社会经济影响、政府救助、危险因素、精神卫生、卫生费用、卫生设施分布、交通运输变化、网络安全、海外人口登记、文献与科研项目统计等方面。前 5 个主题数据集数量占总数 84%（205 个），分别为人口统计学（44.4%）、经济学（16.9%）、政府补助（12.4%）、救济工作（7.0%）和卫生设施（3.3%）。从各国开放数据主题分布看，5 个国家均开放有新冠肺炎病例人口统计特征的数据，如地区分布、年龄分布、性别分布、死亡统计等。不同国家数据主题分布各有特点，如美国数据主题最为丰富，涉及病例统计、经济学、救助与补助、卫生设施、交通运输、体育运动、病毒检测、图书馆服务等 10 余个方面。印度数据仅涉及病例统计 1 个主题。澳大利亚数据以凸显地方政府救济工作数据集为主，其次为病例统计类。加拿大数据更聚焦于疫情对经济影响，其次为

病例统计、政府补助、海外人口登记、交通运输等。英国数据涉及政府补助、经济影响、卫生服务、患病危险因素、行政管理审计等主题。各国开放的新冠肺炎疫情相关数据集中包含多项新冠肺炎专项计划或调查数据，如美国国家卫生统计中心“家庭脉搏调查”^[15]关于疫情对家庭生活影响的数据，美国中小企业管理局“冠状病毒纾困计划”^[16]关于薪资保护、经济损失灾难贷款计划等的数 据，澳大利亚新南威尔士州法律援助项目^[17]关于疫情期间家庭暴力、托幼等援助的数据。

3.5 数据开放等级

5 国政府数据开放等级分布，见图 3。达到 3 星及以上的占多数（180 个，74.1%），以非专用格式的结构化数据形式或使用 URI 标识数据形式开放，代表性格式有 CSV、JSON、XML、RDF 等。美国数据开放等级在 0~4 星均有分布，且仅美国有 0 星（有元数据但无数据文件）和 4 星（有 URI 标识）数据，分别占该国数据 8% 和 66.7%。印度和加拿大数据开放等级分布较为集中，全部或绝大多数为 3 星。澳大利亚和英国数据以 1 星（需专用软件的结构化数据）和 3 星居多，主要为 PDF 和 CSV 格式。值得关注的是美国和印度部分数据提供针对不同用户和使用需求的多种数据下载格式；美、英、澳 3 国部分以非机器可读格式（如 PDF、JPEG）提供的数据并不符合严格意义的政府开放数据要求^[18]。

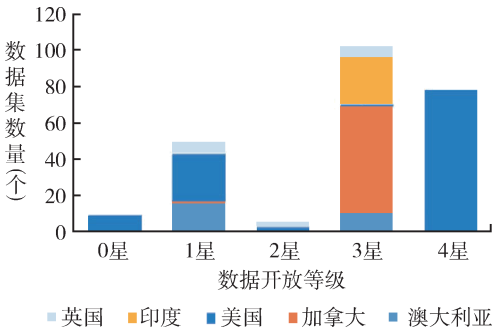


图 3 5 国政府新冠肺炎疫情相关数据集开放等级分布

3.6 开放许可类型

开放许可是政府开放数据的重要步骤和基础保

障。5 国政府开放的 243 个数据集中，68.7%（167 个）标记和描述了许可类型，见表 3。开放许可类型可归纳为开放政府许可（Open Government License, OGL）、国际组织开放许可、发布机构数据使用许可和相关法律法规 4 种。加拿大和印度数据集全部使用本国制定的 OGL，即 OGL Canada^[19] 和印度国家数据共享和获取政策^[20]（National Data Sharing and Accessibility Policy, NDSAP）；英国 OGL^[21] 是中央及地方政府默认和优先使用的许可，但仅有两个数据集（16.7%）标记了该许可，其余 85.7% 数据未做出使用许可说明；澳大利亚未专门制定 OGL，要求除法律规定或有特殊说明外，政府开放数据使用遵循知识共享许可协议（Creative Commons）的 CC BY 3.0 AU^[22]。但该国只有 40% 数据标记了 CC 许可，其余 60% 未做许可说明；美国许可类型最为多样，有版权法关于（联邦）政府作品版权声明^[23]（41%，48 个）、国际开放知识许可协议 CC - PDM、开放数据共用许可 ODC - PDDL^[24]、数据发布机构许可声明几种，也有 41.9% 数据未做许可说明。许可协议是政府开放数据的特点之一^[25]。严格来说，没有明确开放许可的数据甚至不能称为开放数据。美、印、英、加、澳 5 国均针对政府开放数据使用许可的协议或可援引的法律法规，但从新冠肺炎疫情数据开放实践来看，仅加拿大和印度开放许可得到全面执行，澳大利亚差强人意，英国最不理想。

表 3 5 国政府新冠肺炎疫情相关数据集开放许可类型						
协议名称	澳大利亚	加拿大	印度	美国	英国	总计
CC 协议	10	-	-	1	-	11
加拿大 OGL	-	61	-	-	-	61
印度 NDSAP	-	-	26	-	-	26
美国政府作品制度	-	-	-	48	-	48
NIST 数据许可说明	-	-	-	2	-	2
ODC 协议	-	-	-	17	-	17
英国 OGL	-	-	-	-	2	2
无	15	-	-	49	12	76
总计	25	61	26	117	14	243

4 结语

研究发现 5 国政府均能在本国疫情早期及时开放疫情相关数据；汇集的数据来源除卫生部门外包括经济贸易、交通运输、国家统计局等部门；数据主题涉及病例统计、疫情对社会经济影响、民生与救济、大众休闲及运动等领域；多数数据开放等级较高，且对开放使用许可做出规定。国外政府新冠肺炎疫情相关数据开放主要存在以下两方面问题：一是政府开放数据质量管理不规范。5 国政府开放的疫情相关数据中普遍存在数据集描述、数据文件等核心元数据项缺失的现象，而以 PDF、JPEG 等格式提供的数据集使平均可机读率仅为 74.1%，种种数据质量问题将极大影响开放数据利用。二是政府数据开放使用许可贯彻不到位。近 30% 数据未注明任何使用许可类型，以英国和澳大利亚最为明显，而标注使用许可的数据中有 40% 采用政府规定之外的其他许可。在权责声明不够清晰的情况下潜在的法律担忧可能会阻碍政府开放数据使用意愿。多源数据融合应用时，纷杂的授权协议容易造成混乱。由于海外疫情持续发酵，美国、印度、英国、加拿大和澳大利亚 5 个国家政府门户开放疫情相关数据激增，导致前期选取样本代表性减弱，是本研究的主要不足。

参考文献

1 刘瑾. 新型冠状病毒肺炎疫情下地方政府数据开放平台建设进展研究 [J]. 数字图书馆论坛, 2020 (2): 3-9.

2 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.

3 翁士洪, 林晨晖, 早克然·库地热提. 突发事件政府数据开放质量评估研究: 新冠病毒疫情的全国样本实证分析 [J]. 电子政务, 2020 (5): 2-13.

4 The Economic Times. WHO Warns Covid - 19 Entering 'New and Dangerous Phase' [EB/OL]. [2020-07-18]. <https://economictimes.indiatimes.com/news/international/world-news/who-warns-covid-19-entering-new-and-dangerous-phase-says-pandemic-accelerating>

- with - over - 150k - cases - in - a - day/videoshow/76475686. cms.
- 5 World Health Organization. WHO Coronavirus Disease (COVID - 19) Dashboard [EB/OL]. [2020 - 07 - 22]. <https://covid19.who.int/>.
 - 6 韦忻伶, 安小米, 李雪梅, 等. 开放政府数据评估体系述评: 特点分析 [J]. 图书情报工作, 2017, 61 (18): 119 - 127.
 - 7 World Bank Group. Open Data Readiness Assessment Methodology (version 3.1) [EB/OL]. [2020 - 07 - 25]. http://opendatatoolkit.worldbank.org/docs/odra/odra_v3.1_methodology-en.pdf.
 - 8 OECD. Open Government Data [EB/OL]. [2020 - 07 - 27]. <https://www.oecd.org/gov/digital-government/open-government-data.htm#:~:text=Open%20Government%20Data%20%28OGD%29%20is%20a%20philosophy-%20and, and%20commission%20huge%20quantities%20of%20data%20and%20information>.
 - 9 World Wide Web Foundation. Open Data Barometer - Leaders Edition [M]. Washington DC: World Wide Web Foundation, 2018.
 - 10 中国医学科学院医学信息研究所. 中文医学主题词表 [EB/OL]. [2020 - 08 - 12]. <http://cmesh.imicams.ac.cn/index.action?action=index&noMsg=1>.
 - 11 Tim Berners Lee. 5 - star Open Data [EB/OL]. [2020 - 08 - 15]. <http://5stardata.info/en/>.
 - 12 Data. Gov. Washington State Novel Coronavirus (COVID - 19) Cases [EB/OL]. [2020 - 08 - 12]. <https://catalog.data.gov/dataset/washington-state-novel-coronavirus-covid-19-cases>.
 - 13 Ministry of Health and Family Welfare, Government of India. Ministry of Health and Family Welfare [EB/OL]. [2020 - 08 - 22]. <https://www.mohfw.gov.in/>.
 - 14 李雪峰. 健全国家突发公共卫生事件应急管理体系的对策研究 [J]. 行政管理改革, 2020 (4): 13 - 21.
 - 15 U. S. Centers for Disease Control and Prevention. Household Pulse Survey [EB/OL]. [2020 - 08 - 22]. <https://www.cdc.gov/nchs/covid19/health-care-access-and-mental-health.htm>.
 - 16 U. S. Small Business Administration. Coronavirus Relief Options [EB/OL]. [2020 - 08 - 24]. <https://www.sba.gov/funding-programs/loans/coronavirus-relief-options>.
 - 17 Legal Aid New South Wales. Legal Aid NSW [EB/OL]. [2020 - 08 - 25]. <https://www.legalaid.nsw.gov.au/>.
 - 18 Lessig L. "The 8 Principles of Open Government Data" 2007 [EB/OL]. [2020 - 08 - 25]. <https://opengovdata.org/>.
 - 19 Government of Canada. Open Government Licence - Canada [EB/OL]. [2020 - 09 - 01]. <https://open.canada.ca/en/open-government-licence-canada>.
 - 20 Government of India. Government Open Data License - India National Data Sharing and Accessibility Policy [EB/OL]. [2020 - 09 - 01]. <https://data.gov.in/government-open-data-license-india>.
 - 21 The National Archives. Open Government License for Public Sector Information [EB/OL]. [2020 - 09 - 01]. <http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/>.
 - 22 Creative Commons. Attribution 3.0 Australia (CC BY 3.0 AU) [EB/OL]. [2020 - 09 - 02]. <https://creativecommons.org/licenses/by/3.0/au/>.
 - 23 The United States Government. U. S. Government Works [EB/OL]. [2020 - 09 - 02]. <https://www.usa.gov/government-works>.
 - 24 Open Data Commons. Open Data Commons Public Domain Dedication and License (PDDL) v1.0 [EB/OL]. [2020 - 09 - 02]. <https://opendatacommons.org/licenses/pddl/1-0/>.
 - 25 范佳佳. 中国政府数据开放许可协议 (CLOD) 研究 [J]. 中国行政管理, 2019 (1): 23 - 29.