

中国地方政府开放数据研究：技术与法律特性

郑 磊 熊久阳*

【摘要】 论文对我国 13 个地方政府通过平台开放的数据从技术和法律特性上进行了比较研究，包括开放数据的数量、种类、格式、获取方式、及时性、开放授权、元数据、浏览量和下载量等维度。研究发现，各地的开放政府数据集总量偏低，开放数据中满足可机读格式的比例为 66.85%，而满足开放性非专属格式的比例更低（18.78%）。目前，各地承诺实现更新的数据比例为 66.7%，而其中真正按承诺实现更新的数据比例只有 20.83%。许多地方对于开放数据的免费获取、自由利用和传播的权利未予以明确保证。各地开放数据集的元数据所包含的信息种类参差不齐。此外，数据开放的供给和需求也未充分匹配。最后，研究对开放政府数据的标准格式、开放协议、元数据提供、按需导向等方面提出建议。

【关键词】 开放数据 地方政府 评估 中国

【中图分类号】 D63

【文献标识码】 A

【文章编号】 1674 - 2486 (2017) 01 - 0053 - 21

一、引言

近年来，信息技术与经济社会的交汇发展产生和积淀了海量数据，数据已成为国家的基础性战略资源，这其中有大量关键性数据储存在各级政府部门中。开放政府数据，供社会自由免费地获取和利用，有利于提高政府透明度，提升政府治理能力和效率，同时也可以更好地满足社会的数据需求，激发社会创新，促进经济增长。自 2009 年美国联邦数据平台 data.gov 上线后，开放数据的步伐在全球范围内不断发展推进。2015 年 8 月，我国国务院发布了《促进大数据发

* 郑磊，复旦大学国际关系与公共事务学院，副教授；熊久阳，复旦大学国际关系与公共事务学院，硕士研究生。感谢匿名评审人的意见。

基金项目：国家自然科学基金委面上项目“大数据背景下开放政府数据的因素与机理研究：系统动力学建模与政策仿真”（71473048）。

展行动纲要》，提出要在 2018 年建成国家统一数据开放平台。在此之前，我国一些地方政府已先行先试，推出地方性的开放政府数据平台、如北京、上海、武汉、无锡、佛山南海等地。

本文对我国 13 个地方政府通过开放平台提供的数据从技术和法律特性上进行了比较研究。首先，构建了一个开放政府数据评估框架，并基于该框架及其指标体系对十三个地方政府开放数据平台上的数据集数量、种类、格式、获取方式、及时性、开放授权、元数据提供、浏览量、下载量等维度进行了持续性评估，以分析我国地方层面开放数据的进展、特点与问题，并为我国的开放政府数据的发展推进提供建议，以期有助于提升我国政府数据开放水平。

二、开放数据评估现状与不足

随着政府信息化水平的不断发展，政府部门在履行职责过程中生成、采集和保存了海量的数据，成为一个社会主要的数据库。政府所保存的这些数据与公众的生产生活息息相关。在大数据时代，政府向社会公众开放其所持有的数据，供社会进行增值利用和创新应用，为其生产、生活和经济社会活动服务，将创造巨大的公共价值，推动经济增长和社会发展（郑磊，2015a）。

“开放政府数据”将开放的对象推进到了“数据”层，而不仅仅是“信息”层。“数据”是第一手的原始记录，未经加工与解读，不具有明确意义，而“信息”是经过连接、加工或解读之后被赋予了意义的数据库。传统的政府信息公开侧重于保障公众的知情权，而开放政府数据更强调社会利用政府数据的权利，以创造经济与社会价值。因此，开放政府数据是政府信息公开在大数据时代的延伸和跃进（郑磊，2015b）。

开放数据需要满足必要的标准才能带来价值，因此，评估开放数据的数量、标准和质量具有重要意义。目前，主要的国际性开放政府数据评估项目包括万维网基金会（World Wide Web Foundation）和开放数据研究院（Open Data Institute）组织开展的“开放数据晴雨表”（Open Data Barometer）和英国开放知识基金会（Open Knowledge Foundation）组织开展的“全球开放数据指数”（Global Open Data Index）等。夏义堃（2015）对 7 个国际组织开放政府数据评估项目的评估主题、评估侧重点、评估对象和评估方法进行了比较研究。

郑磊、关文雯（2016）通过对 11 个国内外评估项目的评估框架和指标进行深入分析后发现，现有的开放政府数据评估项目主要针对“基础”“平台”“数据”“使用”和“效果”五个层面开展评估。其中，“基础”层是指开放政府数

据所处的内外部环境，这是开放数据的基础和支撑。“平台”层是指开放数据的网上平台，是开放数据的载体。“数据”层是指开放的数据的标准、数量、质量、种类等，这是开放数据的核心。“使用”层是指数据使用的情况，包括使用需求、目的、形式等方面，因为政府开放的数据只有被有效利用，才能真正产生结果。“效果”层是指数据被开放和利用后产生的实际效果与价值。该研究建议在开放数据的不同发展阶段，应因时制宜、循序渐进采用适用于该阶段的评估指标，逐步从针对“基础”这一准备度层面的评估，走向针对“平台”和“数据”等产出层面的评估，并最终开展针对“使用”和“效果”等绩效的评估，不断引导开放数据向纵深推进。

该研究还发现目前开展的评估项目有全球性评估和区域性评估两类，前者如开放数据晴雨表和全球开放数据指数，对全球范围内开放政府数据的国家和地区进行评估；后者如开放政府数据指数、欧洲公共部门信息记分牌、欧洲开放数据监测、开放政府数据可行性研究等，分别对经合组织成员国、欧盟成员国、欧洲国家和部分发展中国家的开放政府数据实践进行评估。

这些国际性或区域性开放数据评估项目主要针对国家层面开展宏观评估，而缺少对一个国家内部不同地方的开放数据实践进行评估。郑磊和高丰（2015）最早对我国地方政府开放数据实践开展了系统评估，并构建了一个包括“基础”“数据”“平台”三个层面13个维度的评估框架。该研究发现我国各地政府的数据开放存在开放数据总量少、可机读比例低、动态数据比例低、数据授权协议不完全符合法律意义上的开放性等问题。然而，该研究并未对数据层面开展更多维度的深入研究，也未能从时间维度持续跟踪我国开放数据的进展。

“数据”层是开放政府数据的核心所在，目前国内外开展的开放数据评估项目全部都对“数据”层面进行了评估（郑磊、关文雯，2016）。本研究将重点对我国地方政府开放数据层面的技术和法律特性开展深入研究。首先基于国内外文献和相关评估实践，并考虑到我国开放数据实际发展现状，构建一个“数据”层评估框架和指标体系，并依此对中国多个地方的开放政府数据实践开展持续性的评估，以客观分析我国地方政府开放政府数据实践的进展、特点与问题。

三、评估框架构建与数据采集

关于开放数据应满足的原则与标准，开放政府工作组（Open Government

Working Group, 2007) 提出了八项关于公共数据的获取和使用原则, 既政府数据在满足以下条件时可称为“开放”数据: (1) 完整; (2) 一手; (3) 及时; (4) 可获取; (5) 可机读; (6) 非歧视性; (7) 非专属性; (8) 无需授权。维斯库斯 (Gianluigi Viscusi) 等 (Viscusi et al., 2014) 提出的开放政府数据完成度评估框架包含数据完整性、准确性和及时性三大质量维度。伯纳-李 (Tim Berners-Lee) (Berners-Lee, 2014) 对开放数据提出五星发展评估模式: 一星: 使用者基于开放授权可在网络上获取数据, 但数据可能采取 PDF、JPEG 等格式; 二星: 使用者可通过可机读方式获取结构化数据 (如取用 Excel 而非图文件表格); 三星: 使用者可获取非专属开放格式的数据 (如 CSV、XML 等); 四星: 使用者可以通过 URI 链接找到每一个数据集; 五星: 数据之间实现关联。2015 年, 来自政府、多边组织、公民社会和私有部门的专家共同制定的国际开放数据宪章将开放数据界定为具备必要的技术和法律特性, 从而能被任何人、在任何时间和任何地点进行自由利用、再利用和分发的电子数据, 并提出六项原则: 默认开放、及时和全面、可获取和可利用、可比较和互操作、为了改善治理与公众参与和为了包容性发展与创新 (Open Data Charter, 2015)。根据世界银行的界定, 能被任何人出于任何目的不受限制进行自由利用、再利用和分发的数据即被视为开放数据。所谓“开放”, 应具备两个维度的特性: 一为技术性开放, 即数据应为可机读、非专属性的电子格式, 从而能被任何人利用通用、免费的软件即可获取和利用, 数据还应置于公共服务器上供公众获取, 不设密码和防火墙; 二为法律性开放, 即这些数据属于公有领域, 对于其自由利用设置最低限制 (World Bank, 2017)。

目前国内外开展的开放数据评估项目在“数据”层面重点关注数据数量、数据质量、数据标准、数据管理、数据类别、关键数据集等指标。其中, 数据格式标准、数据质量两个指标受重视程度最高。在具体指标上, 数据的开放性、可机读性、可获取性、非专属性、及时性、元数据提供等方面最受关注 (郑磊、关文雯, 2016)。

基于以上有关开放数据的标准和现有的国内外开放数据评估框架, 结合我国数据开放的实际情况, 本文构建了一个专注于数据层面的开放数据评估框架, 包括数量、种类、格式、获取方式、及时性、开放授权、元数据提供、浏览量、下载量等方面 (参见表 1)。

表 1 研究框架

一级指标	二级指标	描述说明
数据集数量		开放的数据集总数
数据种类		开放的数据集的主题种类
可机读格式		数据以机器可读的格式开放，例如 XLS 格式，而非 PDF 格式
开放格式		数据以开放的、非专属格式提供，不限制任何人使用，确保其无需某个特定的（特别是付费的）应用程序才能访问。例如 CSV 格式，而非 Excel 格式
数据接口		以 API 数据接口方式提供数据
及时性	承诺的更新频率	数据承诺的更新频率：分为“实时更新”“以年/月/周/日更新”“不定期更新”“无更新”“其他”。“其他”类包括未标注更新方式等情况
	按承诺实现更新	数据按照其所承诺的更新频率实现了更新
开放授权	具有数据开放授权协议	所开放的数据具有相应的开放授权协议
	非歧视性	开放授权协议明确保障用户可平等访问、获取、使用和分享数据
	免费获取	开放授权协议明确保障用户可免费访问、获取、使用和分享数据
	自由利用	开放授权协议明确保障用户可自由增值利用和再利用数据
	自由传播分享	开放授权协议明确保障用户可自由传播和分享数据
元数据提供	元数据信息种类	元数据包含的信息种类，如主题分类、摘要、来源单位、更新频率、格式、语种、安全分级、记录条数、文件大小、发布时间、浏览数、下载数等
	元数据集中展示	元数据信息集中显示于数据下载页面，而不是分散存在，需要用户打开不同页面才能查看
浏览/下载量	下载量	开放数据的下载量统计
	浏览量	开放数据的浏览量统计

资料来源：作者自制。

基于以上评估框架，研究采用了自动抓取和人工观察两种方法采集数据，并分别在三个时间段采集了数据。第一批数据采集时间为 2015 年 5 月 20 日，

◆ 专栏：大数据与公共治理

通过公开报道和搜索引擎发现已上线开放数据平台的地区，并综合考虑各地开放数据的成熟水平、行政层级和地域分布，选择了北京市、上海市、武汉市、无锡市、湛江市、宁波市海曙区、佛山市南海区等地方的开放政府数据平台作为研究样本。2016年，随着更多的地方城市推出开放数据平台，并整合调整省市级数据平台，本研究又分别于2016年1月15日和2016年11月26日采集了两批数据，既包括2015年已采集数据的平台，也不断纳入新上线开放数据的平台的地方作为研究样本，包括浙江省、广东省、广州市、湛江市、东莞市、肇庆市、贵州省的开放数据平台，而宁波市海曙区的平台已并入浙江省平台，不再独立运行。因此，本研究共针对13个地方的开放数据平台采集了数据，具体样本列表参见表2。

表2 研究样本

样本对象	行政级别	所处地区	开放数据平台名称	开放数据平台网址
北京市	直辖市	华北地区	北京市政务数据资源网	bjdata.gov.cn
上海市	直辖市	华东地区	上海市政府数据服务网	datashanghai.gov.cn
贵州省	省	西南地区	贵州省政府数据开放平台	gzdata.gov.cn
浙江省	省	华东地区	浙江政务服务网“数据开放”专题网页	data.zjzfwf.gov.cn
广东省	省	华南地区	开放广东—广东省政府数据统一开放平台	gddata.gov.cn
武汉市	副省级市	华中地区	武汉市政府公开数据服务网	wuhandata.gov.cn
青岛市	副省级市	华东地区	青岛政务网—政府数据开放	data.qingdao.gov.cn
广州市	副省级市	华南地区	广州市政府数据统一开放平台	datagz.gov.cn
无锡市	地级市	华东地区	无锡市政府数据服务网	opendata.wuxi.gov.cn
湛江市	地级市	华南地区	湛江数据服务网	data.zhanjiang.gov.cn
东莞市	地级市	华南地区	数据东莞	dataopen.dg.gov.cn
肇庆市	地级市	华南地区	肇庆市人民政府网—用数据	zhaoqing.gov.cn/sjkh/
佛山市南海区	地级市 下辖区	华南地区	数说南海	data.nanhai.gov.cn

资料来源：作者自制。

四、研究发现

(一) 数据集开放总量与可机读数量比较

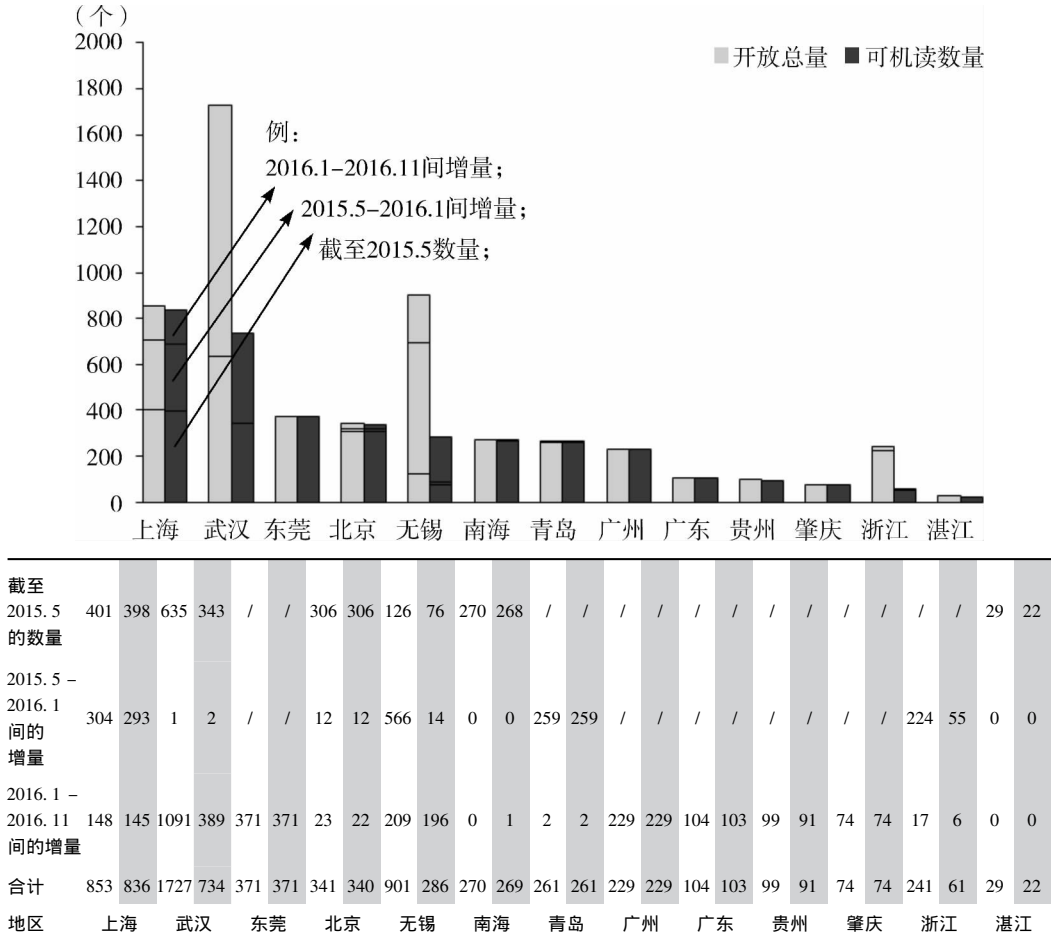


图1 开放数据总量与可机读数量对比

资料来源：作者自制。

截至2016年11月26日，在研究样本所覆盖的各地开放政府数据平台上，武汉市、无锡市、上海市的开放总量相对最高，将三个时间点采集的数据量比较后发现，这三个城市的数据集开放总量在2016年增量都有较大增长，而其他地区的数据集量增长相对缓慢甚至停滞，湛江市、佛山南海区的开放数据集总量最近一年几乎没有变化。2016年新推出开放数据平台的东莞市、广东省、广州市、贵州省、肇庆市的数据量级多在一百到三百左右。需要特别说明的是，

◆专栏：大数据与公共治理

青岛市 2016 年下半年新增的数据集三千余个，总数达到 3706 个，但其中大量“数据集”中仅有 1 条或者 0 条数据（共 3445 个），这些单条和 0 条数据并不是“数据集”，剔除这类数据后，青岛市目前实际开放的数据集总数为 261 个。

为方便用户获取和利用数据，数据集应以可机读格式开放（例如 XLS，而非 PDF 格式）。以可机读格式的标准来看（图 1 中右侧的深色条柱），上海市和武汉市的可机读数据总量最高，分别为 836 个和 734 个，东莞市、北京市、无锡市等地区紧随其后。各地方平均有 66.85% 的公开数据符合可机读格式的要求，整体比例虽然相对 2015 年 5 月的情况（81.1%）（郑磊、高丰，2015）有所下降，但大部分单个地区的可机读比例已较高，东莞市、青岛市、广州市、肇庆市 4 个地区开放的数据集符合可机读格式的比例最高，全部达到 100%；北京市、佛山市南海区、广东省 3 个地区的比例接近 100%；上海市、贵州省的可机读格式也都超过了 90%。尤其是 2016 年新上线的开放数据平台如东莞市、广东省、广州市、贵州省、肇庆市等在可机读比例方面表现都较好，体现了这些地方已普遍认识到开放数据需满足可机读格式的要求。然而，目前武汉市、无锡市和浙江省的可机读比例仍然较低，其中武汉市有超过四分之一的数据为 PDF 格式，无锡市有过半数据集是网址链接形式，浙江省的数据集格式较为杂乱且大部分为不可机读格式（参见表 3）。

表 3 可机读格式比例

指标/ 地区	东莞	青岛	广州	肇庆	北京	南海	广东	上海	贵州	湛江	武汉	无锡	浙江
开放数据 集总量	371	261	229	74	341	270	104	853	99	29	1 727	901	241
可机读数 量占比	100%	100%	100%	100%	99.71%	99.63%	99.04%	98.01%	91.92%	75.86%	42.50%	31.74%	25.31%

资料来源：作者自制。

(二) 开放数据集的主题种类 表 4 开放数据集的主题种类分布表

主题领域	北京	上海	武汉	无锡	青岛	南海	浙江	湛江	贵州	广东	广州	肇庆	东莞	总计	
经贸金融		经济建设 222 金融发展 450 金融服务 41			经济发展 111 财税金融 10		经济建设 73		经济建设 15 民生服务 13 民生服务 10 企业服务 56	经济建设 15 民生服务 10 企业服务 56 财税金融 6	经济发展 15 民生服务 10 企业服务 56	经济建设 11		968	
民生服务	生活服务 37 企业服务 46	民生服务 69 公共生活服务 326 企业服务 369 社会民生 1	公共生活服务 69 公共生活服务 326 企业服务 369 社会民生 1	生活服务 191 企业服务 369 社会民生 1	生活服务 16 企业服务 48	生活服务 29 企业服务 48	民生服务 23	民生服务 23	民生服务 13 民生服务 10 民生服务 10 民生服务 10	民生服务 13 民生服务 10 民生服务 10 民生服务 10	民生服务 15 民生服务 10 民生服务 10 民生服务 10	民生服务 10		1249	
机构团体	政府机构与 46 社会团体 46	政府机构与 49 社会团体 46	政府机构与 447 社会团体 93	政府机构与 447 社会团体 93	政府机构与 10 社会团体 18	政府机构与 10 社会团体 18	机构团体 32	政府机构 32	机构团体 7 机构团体 11 机构团体 11 机构团体 11	机构团体 11 机构团体 11 机构团体 11 机构团体 11	机构团体 12			725	
医疗卫生	健康 35 文体娱乐 35 文化旅游 24	卫生健康 84 文化娱乐 48 文化旅游 24	医疗卫生 45 文化娱乐 47 文化旅游 15	医疗健康 82 文化娱乐 20 旅游住宿 15	医疗健康 34 文化娱乐 14 旅游住宿 8	医疗健康 28 文化娱乐 18 旅游住宿 8	医疗健康 28	医疗健康 28	医疗健康 12 文化休闲 1 文化休闲 1 文化休闲 1	医疗健康 12 文化休闲 1 文化休闲 1 文化休闲 1	医疗健康 5 文化休闲 13 文化休闲 24 文化休闲 24	卫生健康 34 文化休闲 24 文化休闲 24 文化休闲 24	卫生健康 8		367
食购游娱	文体娱乐 3 餐饮美食 3 消费购物 5	文化娱乐 48 文化旅游 24	文化娱乐 47 文化旅游 15	文化娱乐 20 旅游住宿 15	文化娱乐 14 旅游住宿 8	文化娱乐 18 旅游住宿 8	文化休闲 30	文化休闲 30	文化休闲 1 文化休闲 1 文化休闲 1	文化休闲 1 文化休闲 1 文化休闲 1	文化休闲 13 文化休闲 24 文化休闲 24	文化休闲 5		317	
交通出行	道路服务 27 资源环境 27	道路服务 72 资源环境 47	交通服务 93 资源环境 109	交通服务 8 资源环境 109	道路服务 4 资源环境 7	道路服务 26 资源环境 30	道路交通 16	交通运输 16	交通运输 3 交通运输 3 交通运输 3	道路交通 1 资源环境 11 资源环境 32	道路交通 9 资源环境 16 资源环境 16	道路交通 4		280	
环境资源	资源环境 21 保护 21	资源环境 47 保护 47	资源环境 109 保护 8	资源环境 109 保护 8	资源环境 7 保护 3	资源环境 30	资源环境 30	资源环境 30	资源环境 3 资源环境 3 资源环境 3	资源环境 11 资源环境 32 资源环境 32	资源环境 16 资源环境 16 资源环境 16	资源环境 7		306	
城建住房		城市建设 95	城市建设 95	城市建设 95	城市建设 0 城市管理 0	城市建设 13 城市管理 0	城市建设 25	城市规划 25	城市建设 0 城市建设 12 城市建设 12	城市建设 12 城市建设 12 城市建设 12	城市建设 11 城市建设 11 城市建设 11	城市建设 8		193	
教育科研	教育科研 11 社会保障 11	教育科技 81 社会发展 46	教育科技 99 社会发展 46	教育科研 47 社会保障 1	教育科技 19 社会保障 11	教育科技 18 社会保障 15	教育科技 12	教育科技 12	教育科技 11 教育科技 15 教育科技 15	教育科技 11 教育科技 15 教育科技 15	教育科技 22 教育科技 22 教育科技 22	教育科技 8		332	
社会保障	社会保障 11 劳动保障 11	公共生活安全 29 公共生活安全 29	公共生活安全 20 公共生活安全 20	社会保障 47 劳动保障 1	社会保障 11 劳动保障 3	社会保障 15	社会保障 15	社会保障 15	社会保障 15 社会保障 15 社会保障 15	社会保障 15 社会保障 15 社会保障 15	社会保障 6 社会保障 6 社会保障 6	社会保障 8		120	
公共安全	生活安全 6 公共生活安全 6	公共生活安全 29 公共生活安全 29	公共生活安全 20 公共生活安全 20	公共生活安全 8 公共生活安全 8	公共生活安全 13 公共生活安全 13	公共生活安全 2	公共生活安全 2	公共生活安全 2	公共生活安全 7 公共生活安全 7 公共生活安全 7	公共生活安全 7 公共生活安全 7 公共生活安全 7	公共生活安全 2 公共生活安全 2 公共生活安全 2	公共安全 5		106	
农业农村	农业农村 22 宗教信仰 7	农业农村 14 宗教信仰 7	农业农村 7 宗教信仰 30	农业农村 7 宗教信仰 30	农业农村 2 宗教信仰 6	农业农村 11	农业农村 11	公安 14	公安 14	公共安全 7 公共安全 7 公共安全 7	农业农村 9			65	
宗教司法	宗教信仰 7 许可处罚 7	法律服务 36	宗教信仰 30 法律服务 36	宗教信仰 30 法律服务 36	法律服务 9 许可处罚 0	宗教信仰 6	宗教信仰 6	公安 14	公共安全 7 公共安全 7 公共安全 7	公共安全 7 公共安全 7 公共安全 7	农业农村 9			43	
许可处罚	许可处罚 7	法律服务 36	宗教信仰 30 法律服务 36	宗教信仰 30 法律服务 36	法律服务 9 许可处罚 0	宗教信仰 6	宗教信仰 6	公安 14	公共安全 7 公共安全 7 公共安全 7	公共安全 7 公共安全 7 公共安全 7	农业农村 9			45	
信用服务	信用服务 11	信用服务 11	信用服务 11	信用服务 11	信用服务 11	信用服务 11	信用服务 11	信用服务 11	信用服务 11	信用服务 11	信用服务 11			292	
其他			审计监督 0	审计监督 0	审计监督 0	统计 1	统计 1	统计 1	统计 1	统计 1	统计 1			1	
			许可处罚 277	许可处罚 277	许可处罚 277	行政处罚 15	行政处罚 15	行政处罚 15	行政处罚 15	行政处罚 15	行政处罚 15			277	
			失信曝光 40	失信曝光 40	失信曝光 40	失信曝光 40	失信曝光 40	失信曝光 40	失信曝光 40	失信曝光 40	失信曝光 40			40	
			优良信誉 39	优良信誉 39	优良信誉 39	优良信誉 39	优良信誉 39	优良信誉 39	优良信誉 39	优良信誉 39	优良信誉 39			39	

资料来源：作者自制。

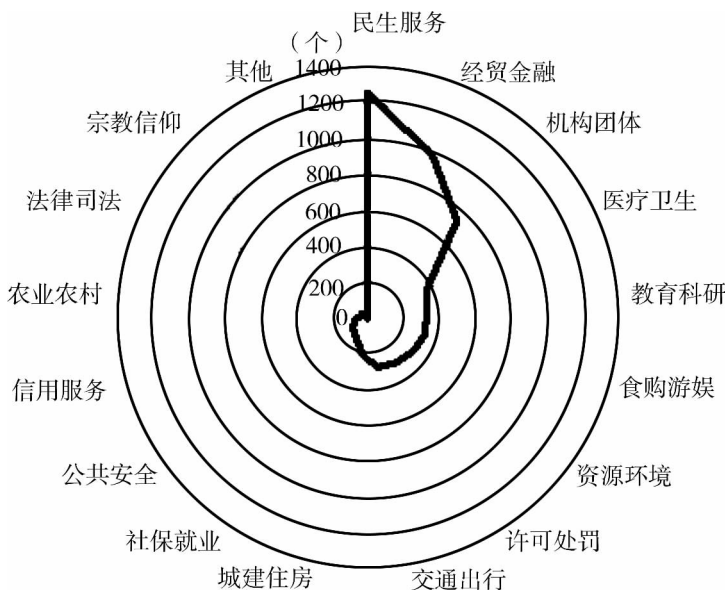


图2 开放数据集的主题种类分布图

资料来源：作者自制。

研究对各个地方开放数据集的主题种类进行了归类梳理（参见表4），各地区的主题分类方法各有差异，主题数量从4（东莞市）到19（青岛市）不等，本文将其梳理为经贸金融、民生服务、机构团体、医疗卫生等17个大类。

分析表明，目前各地开放的数据主要集中于民生服务、经贸发展、机构团体三大类，不同主题种类的分布很不均匀，且地区间差异显著。某些主题种类的命名方式更多反映了部门条线思维的影响，如经济建设、社会保障、交通运输等种类的用词过于正式和官方，与数据利用者和普通公众的用词习惯存在差异，不利于其发现和利用所需数据。有些地区在某些设立的主题分类下暂无实际开放的数据集，比如广东省的“民生服务”“社会资源”等4个主题。信用数据自2016年开始出现在各地区的开放数据主题分类中，比如上海市、贵州省和肇庆市。以上这些地区间差异可能与各地发展特色、开放意愿、数据储备基础等因素有关。

（三）开放格式

开放数据应以开放的、非专属的格式提供，任何实体不得排除他人使用数据的权力，以确保数据无需某个特定的（特别是付费的）应用程序才能访问。例如，CSV属于开放格式，而XLS则不是。研究发现，各样本地区开放的数据集中，满足开放格式的数据平均比例为18.78%。目前，北京市和东莞市100%

提供了开放格式，佛山市南海区 98% 的数据符合开放格式，贵州省和上海市也有小部分数据提供了开放格式，而青岛市、武汉市等 7 个地区的数据以 EXCEL、XLS 等格式居多，虽然是可机读格式，但并非开放格式（参见表 5）。目前，佛山市南海区同时提供了多种格式（CSV、XLS、TXT、XML 和 JSON），东莞市也同时提供了 2 种格式（XLS、JSON），以更好地方便不同需求的数据使用者获取和利用数据，湛江市未提供数据的格式信息。

表 5 数据格式开放性统计

指标/地区	东莞	北京	南海	贵州	上海	青岛	武汉	无锡	浙江	广州	广东	肇庆	湛江
开放总量	371	341	270	99	853	261	1 727	901	241	229	104	74	29
数据提供在开放格式下的比例	100%	100%	98%	9%	6%	0	0	0	0	0	0	0	/

资料来源：作者自制。

（四）数据接口

除数据下载外，有些地区的网站还提供了应用程序编程接口（API 接口，Application Programming Interface）。与下载方式获取数据相比，接口使数据使用者可通过参数有选择性地获取所需的数据用于特定的使用目的，以实时高效满足不同应用场景下的数据需要。北京市、武汉市为所有数据集提供统一接口，东莞市、浙江省、青岛市、上海市、广东省、广州市、贵州省为单个数据集提供单一接口，其中东莞市为每个数据集都提供了接口，总量达到 371 个。其他 4 个地区尚未提供 API 接口（参见表 6，“√”表示“是”、“×”表示“否”）。

表 6 数据接口

指标/地区	北京	武汉	东莞	浙江	青岛	上海	广东	广州	贵州	无锡	肇庆	湛江	南海
是否开放 API 接口	√	√	√	√	√	√	√	√	√	×	×	×	×
接口数量	统一接口	统一接口	371	135	129	102	16	10	3	0	0	0	0

资料来源：作者自制。

(五) 数据及时性

1. 承诺的更新频率分布

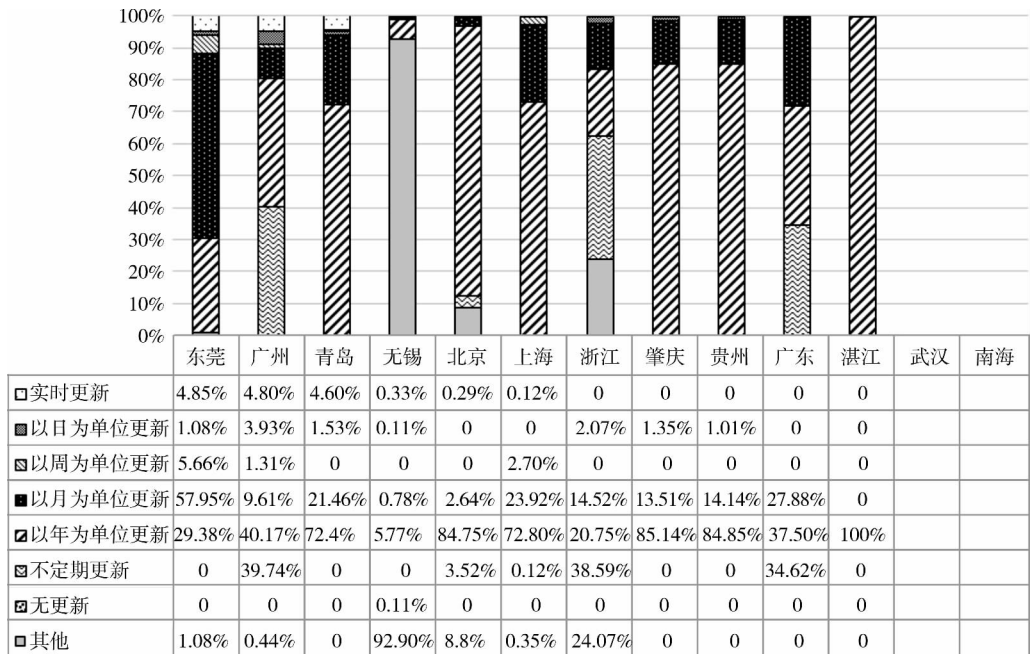


图3 各地区承诺的更新频率

资料来源：作者自制。

研究对各地方平台所承诺的更新频率进行了分类统计（图3）。分为实时更新、以年/月/周/日更新、不定期更新、无更新和其他等类别。

通过对不同时间段所采集的数据进行比较分析，截止到2016年11月，各地整体上承诺以年及年以下为单位进行更新的数据比例为66.7%，20.53%的数据承诺为相对动态的数据（以月、周、日为单位更新或实时更新），79.47%的承诺为相对静态的数据（以年为单位更新、不定期更新、无更新及其他）（参见图4）。而在2015年5月采集的数据中，有13.75%的数据为承诺相对动态更新的数据，86.25%的数据属于承诺相对静态更新的数据（郑磊、高丰，2015），2016年略有进步。在各地样本中，承诺为相对动态数据比例最高的城市为东莞（69.54%），最低为湛江（0%）。武汉市、佛山市南海区两个地区未公布承诺的更新频率，因此该项数据为空白。无锡也有92.90%的数据未注明更新周期，归入“其他”类。

2. 按承诺实现更新的比例

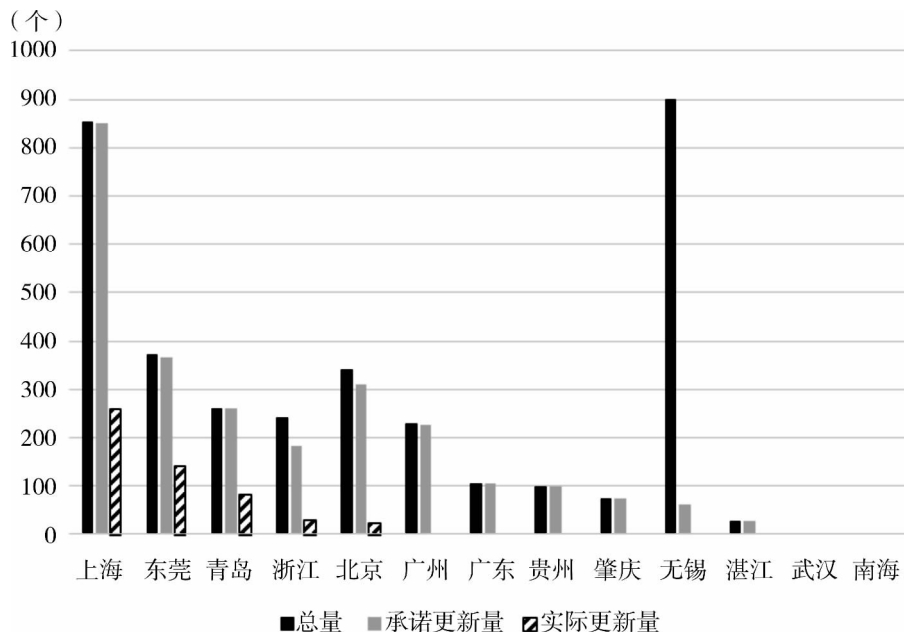


图4 按承诺更新量与比例情况

资料来源：作者自制。

本研究根据各平台上数据集的发布时间、实际更新时间和承诺的更新频率来判定其是否按照其承诺实现了更新。整体而言，2016年11月与2015年5月相比，在承诺更新的数据集之中（即承诺以年及年以下单位进行更新），实际实现更新的数据比例有所增长，从17.21%增加到20.83%。截止到2016年11月，上海市在各地区中数据集实际更新量最多（259个），比例达到30.47%，而东莞市在各地方中按承诺实现更新比例最高（38.69%），其次为青岛市（31.03%）、浙江省（15.85%）、北京市（7.72%），其他城市或未曾更新，或因平台新近上线尚未更新数据。武汉市和佛山市南海区未提供更新频率的信息，故无法统计按承诺更新情况（见图5）。

由于数据的更新频率需有足够可观察周期，部分数据集还无法判断其按承诺更新情况，例如对于更新频率以年为更新频率的数据集，如为最近一年内发布的，目前还无法判定此类数据是否按承诺实现了更新。此外，承诺为不定期更新的数据集，也无法判断其是否已按要求更新，故也不在可观察范围内。

从分析结果看，此次研究中因数据更新信息缺失而无法观察到的数据量占很大比例（见图5）。如图4，武汉市和佛山市南海区完全未提供有关更新频率的信息，无锡市、贵州省、广州市、肇庆市4个地区未提供数据集的更新时间

◆专栏：大数据与公共治理

或发布时间信息。更新频率、发布日期、更新日期三项信息中只要缺失一个，即无法计算其实际更新频率，图5展示了各地由于信息缺失而无法统计实际更新频率的情况（参见图5）。

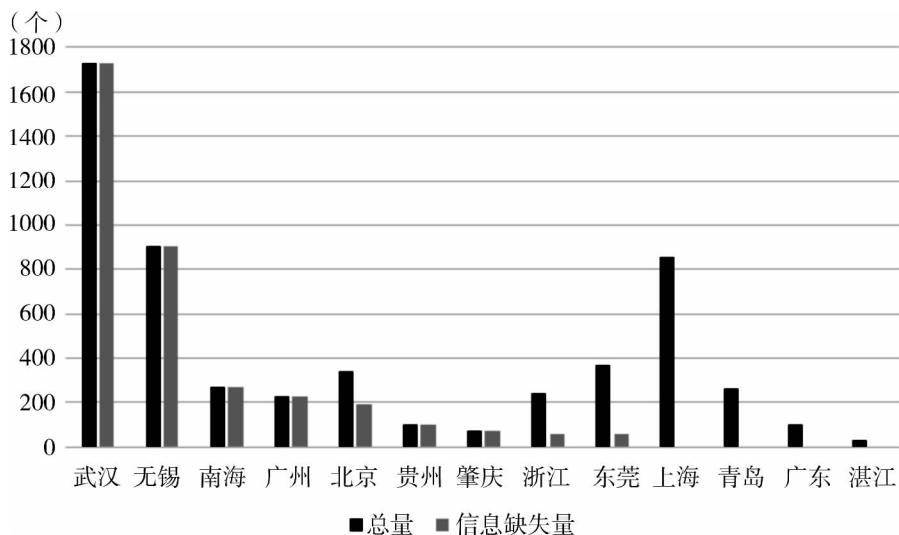


图5 数据更新相关信息缺失情况

资料来源：作者自制。

（六）开放授权

1. 数据开放授权协议

除了通过可机读和开放格式从技术上保证数据集的开放性，还需要开放授权协议从法律上保障数据的开放性。数据授权协议通常包含在开放数据平台的免责条款或用户协议中。研究发现，目前各地平台中仅浙江省和肇庆市的开放政府数据平台上缺少专门的数据开放协议。浙江省的开放数据平台是浙江政务服务网的一个板块，肇庆市的“用数据”也是肇庆市政府门户网站的一个版块，这两个网站的声明并非专为开放政府数据板块准备，也没有专门涉及数据开放授权的内容。各地总体情况见表7（“√”表示有数据开放授权协议、“×”表示没有协议）。

表7 数据开放授权协议

指标/地区	北京	上海	贵州	广东	武汉	青岛	广州	无锡	湛江	东莞	南海	肇庆	浙江
开放授权协议	√	√	√	√	√	√	√	√	√	√	√	×	×

资料来源：作者自制。

2. 开放数据的非歧视性

非歧视性是指平台所发布的数据明确保障任何人可平等地访问、获取、使用和分享。研究发现，目前各地方的开放数据平台都能保障数据开放的非歧视性，对任何用户都予以相同的数据访问和获取权限（参见表8，“√”表示保障了数据开放的非歧视性、“×”表示没有保障）。

表8 开放数据的非歧视性

指标/地区	北京	上海	浙江	贵州	广东	武汉	青岛	广州	无锡	湛江	肇庆	东莞	南海
非歧视性	√	√	√	√	√	√	√	√	√	√	√	√	√

资料来源：作者自制。

3. 开放数据的免费获取

政府数据是行政机关在履行职责过程中制作或者获取的，属于公共资源，在不违反相关政策法规和不损害公共利益的情况下，应免费向社会开放，以供其获取和利用（郑磊，2015a）。目前，仅东莞市未对数据免费时间设置期限，其他9个地区则在授权协议中设有一个模糊的数据免费期限，却并未明确说明是否永久免费，例如“当前免费，但保留收费的权利”和“现阶段免费”等表述。而在无锡市政府数据服务网上，在常见问题栏目中，对于“使用无锡市政府数据服务网上的数据和服务需要收费吗？”这一问题，回答为“无锡市政府数据服务网中，政府部门提供的数据产品目前都是免费的，保留收费的权利”；而在网站的免责声明里则表述为“你有权免费获取本网站所提供的所有政务数据资源”，虽然在后者的声明中明确了免费原则，但在前者问答中声明中又保留了收费的权利，前后表述不一致。浙江和肇庆未提及（见表9）。

表9 开放数据的免费获取

指标/地区	东莞	北京	上海	贵州	广东	武汉	青岛	广州	湛江	南海	浙江	肇庆	无锡
免费开放	是	未明确	未明确	未明确	未明确	未明确	未明确	未明确	未明确	未明确	未提及	未提及	表述不一致

资料来源：作者自制。

4. 开放数据的自由增值利用

开放数据时应明确允许使用者可不受限制地对数据进行商业和非商业性的增值利用。研究发现，目前这一点并未在各地的授权条款中得到明确体现，仅青岛市明确指出可以“基于本网站提供的数据进行商业或非商业性的开发活

◆专栏：大数据与公共治理

动”，肇庆市指出用于商业时需征得相关许可，其他 11 个地区未做说明（见表 10）。

表 10 开放数据的自由利用

指标/地区	青岛	北京	上海	贵州	浙江	广东	武汉	广州	无锡	湛江	东莞	南海	肇庆
自由利用	允许	未提及	未提及	未提及	未提及	未提及	未提及	未提及	未提及	未提及	未提及	未提及	不允许

资料来源：作者自制。

5. 开放数据的自由转播

对于数据的自由传播分享权利，大部分地方包含了不得有偿或无偿转让在本网站中获取的数据资源。其中，青岛市只限制了有偿转让从本网站获取的数据，而其他城市则同时限制了有偿和无偿转让。对于免费的开放数据确实应限制其有偿转让，但限制用户无偿传播和分享开放数据并不符合开放数据的原则，在实际操作中也很难真正限制（见表 11）。

表 11 开放数据的自由传播

指标/地区	青岛	北京	广东	武汉	广州	无锡	南海	上海	贵州	浙江	湛江	东莞	肇庆
自由传播分享	限制有偿转让	限制	限制	限制	限制	限制	限制	未提及	未提及	未提及	未提及	未提及	未提及

资料来源：作者自制。

（七）元数据提供

对于开放的数据提供元数据信息有助于数据利用者清晰地了解数据集的内容与背景，从而更好地利用数据。目前，各地开放的数据集所提供的元数据信息种类数量不一，最高的为北京（17），最低的贵州（6），具体情况见表 12（“√”表示元数据包含该字段、“×”表示不包含）。在元数据显示方式中，有些地方以图标而非文字的形式展现，不利用数据获取，如浙江省、广东省、武汉市、青岛市、广州市、肇庆市、东莞市和佛山南海区等地将数据格式、大小以图标方式显示，这种情况在表 12 中标为三角形（“△”）。

表 12 元数据信息

指标/地区	北京	上海	浙江	贵州	广东	武汉	青岛	广州	无锡	湛江	东莞	肇庆	南海
主题/分类	√	√	×	√	√	√	√	√	√	√	√	√	√
摘要/介绍	√	√	√	√	√	√	√	√	√	√	√	√	√
来源/单位	√	√	√	√	√	√	√	√	√	√	√	√	√
来源机构简介	×	×	×	×	×	√	×	×	×	×	×	×	×
来源机构地址/联系方式	×	√	×	×	×	√	×	×	√	√	×	√	×
关键字	√	√	√	×	×	√	√	×	√	√	√	√	√
更新周期	√	√	√	√	√	×	√	√	√	√	√	√	×
格式	√	√	△	√	△	△	△	△	√	×	△	△	√
语种	√	×	×	×	×	×	√	×	×	×	×	×	×
安全分级/公开属性	√	√	×	×	×	×	×	×	×	√	×	√	×
资源记录数/数据量	√	×	×	×	√	√	√	√	×	×	√	×	√
文件大小	√	×	√	×	×	×	×	×	√	×	×	×	△
数据访问网址	√	×	√	×	×	√	×	×	√	√	×	√	√
浏览次数	×	√	×	×	√	√	√	√	√	√	√	×	×
下载次数	√	√	√	×	√	√	√	√	√	√	√	×	√
上传时间	√	×	√	×	×	×	×	×	×	×	×	×	√
发布时间	√	√	√	×	√	√	√	√	×	√	√	√	×
更新时间	√	√	√	√	×	√	√	×	×	√	√	×	√
评分/评论	√	√	×	×	√	√	√	√	×	√	×	×	√
API 接口信息	√	√	×	×	×	×	×	×	×	×	√	×	×
资源状态	×	×	×	×	×	√	×	×	×	×	×	×	×
适用范围	×	×	×	×	×	×	×	×	×	×	×	×	×
小计	17	14	10	6	9	14	12	9	11	13	11	10	11

资料来源：作者自制。

元数据信息应集中展示便于用户一次性读取，而不需要打开不同页面才能查看。除贵州、广州、肇庆外，大部分地区未将元数据信息集中显示。数据更新时间这一信息分散显示的情况尤为普遍（比如浙江、上海、湛江、北京），其次为下载量或浏览量（比如武汉和无锡），如广东、南海、青岛、东莞也都有各自分散显示的信息，其中南海区是将部分字段做为“数据说明”显示为一个新的页面（参见表 13，“√”表示“集中展示”、“×”表示“分散展示”）。

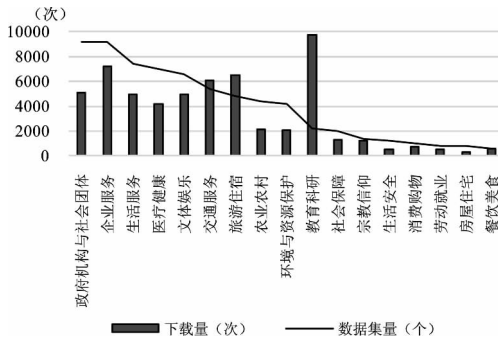
表 13 元数据集中展示

指标/地区	贵州	广州	肇庆	浙江	北京	上海	广东	武汉	青岛	无锡	湛江	东莞	南海
元数据集中展示	√	√	√	×	×	×	×	×	×	×	×	×	×

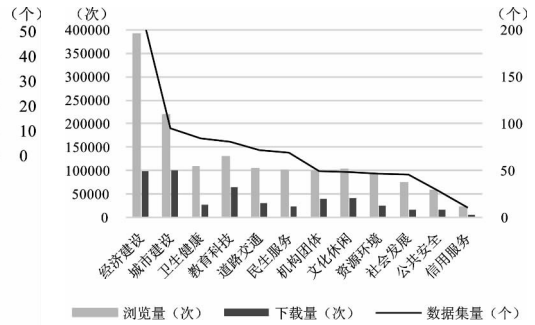
资料来源：作者自制。

(八) 浏览/下载量

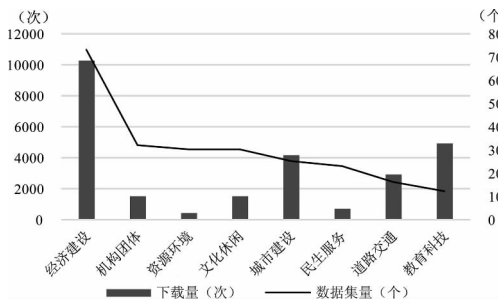
北京市开放数据集统计情况



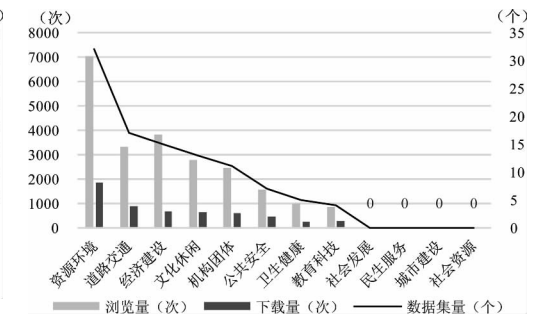
上海市开放数据集统计情况



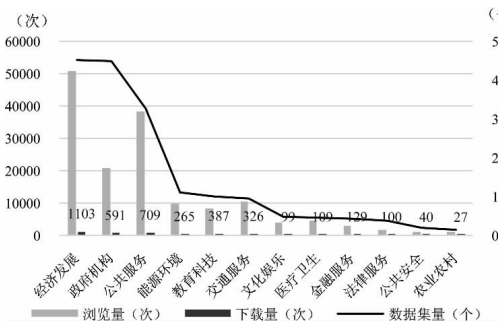
浙江省开放数据集统计情况



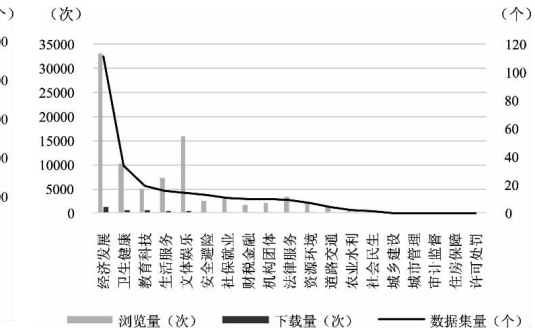
广东省开放数据集统计情况



武汉市开放数据集统计情况



青岛市开放数据集统计情况



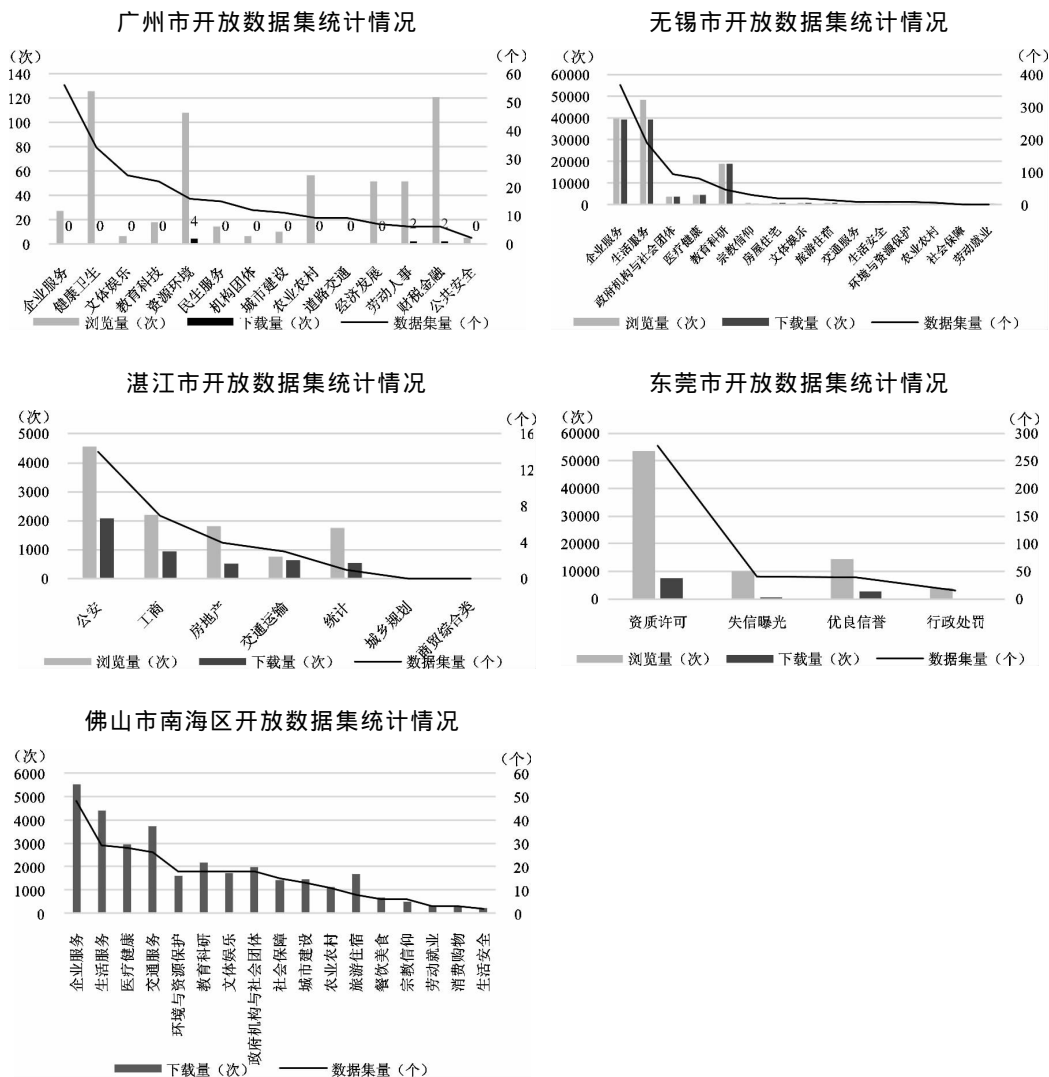


图6 各地开放数据的浏览量/下载量

资料来源：作者自制。

数据的浏览量/下载量（见图6）反映了数据利用用户对数据的兴趣和获取情况，上海市、广东省、武汉市、青岛市、广州市、无锡市、湛江市、东莞市8个地区提供了这两项信息，其他地区则未完全提供该项信息，如北京市、佛山市南海区没有提供浏览量信息，而贵州省、肇庆市则浏览量和下载量信息都没有提供，因而无法进行统计分析。

截至2016年11月，从数据种类上看，与经商与民生相关主题的用户浏览量和下载量相对较高。从浏览量数量级来看，上海市、浙江省、武汉市、青岛市、无锡市、东莞市均超过万次，其中上海已突破10万级，这与各平台开放数据的

数量、价值和时间有关。将浏览量和下载量两项数据的对比分析后发现，下载量普遍小于浏览量，即“看得多，下的少”，原因可能是数据量偏小，可以“一目了然”，因而不需要下载；或由于数据价值不高，没必要下载；或由于下载数据不便捷。目前，武汉市、青岛市、广州市的下载量都远远低于浏览量。

将各主题数量集的数量从高到低排序后可见，上海市、广东省、武汉市、湛江市、东莞市、佛山市南海区6个地区各主题下的数据浏览量和下载量与其数量集的数量基本成正比，即数据集开放数量越大的主题，其浏览量和下载量也越高。青岛市和无锡市的趋势也基本类似，但有个别数据类别出现浏览量或下载量小高峰。北京市和浙江省的下载/浏览量峰值出现在所含数据集数量靠后的主题，虽然这些主题开放的数据集不多，但用户的兴趣却相当高。广州市的浏览量/下载量的变化与主题类型之间未显示出相关性，可能与该平台建立时间较短，目前下载量很少有关。

五、总结与启示

研究发现，我们各地的开放政府数据近年来已取得一定突破和进展，但整体情况仍处于发展初期，有待进一步提升和改进。总体上，各地政府开放的数据集总量偏低，地区间差异较大。开放的数据主要集中在民生服务、经贸发展和机构团体三大类别。数据主题分类的方法和名称仍更多反映了政府部门的视角，而非用户视角，不利用用户识别和发现数据。

在开放数据的技术特性上，目前各地开放的数据中满足可机读格式的比例还不高（66.85%），而满足开放性非专属格式的比例更低（18.78%）。在数据获取方式上，除数据下载外，不少地区已开始通过API接口提供数据。在数据及时性上，承诺实现更新的数据比例偏低（66.7%），而其中真正按承诺实现更新的数据比例只有20.83%，少数地方新增数据集增长缓慢或陷入停滞状态。

在开放数据的法律特性上，部分地区还未提供数据开放授权协议，各地的条款虽都能确保开放数据的非歧视性，但对于开放数据的免费性和自由利用的条款未明确保证或未提供相关信息。而对于开放数据的自由传播的权利，则普遍设有限制或未予以明确。

关于开放数据集的元数据提供，各地情况参差不齐，但未能做到集中展示，展示形式也不统一。此外，许多地方的开放数据平台缺少更新频率、发布日期、更新日期、开放授权协议、浏览量、下载量等信息，不利于用户理解、获取和利用数据。研究还发现数据开放的供给和需求并未完全匹配，开放数据最多的

主题类别并不一定是数据浏览量下载量最大的类别。

基于以上发现，本研究对于我们各地的开放数据实践提出以下建议。首先，开放数据应具备必要的技术性特征，满足可获取、可机读、非专属、及时性等标准。其次，开放数据应具备必要的法律性特征，提供充分的数据开放授权，保障数据获取和利用的非歧视性、开放数据的免费性、对开放数据进行增值利用和传播的权利等。第三、为开放数据提供集中、详实、格式标准的元数据，以降低用户在获取和利用数据时的门槛和限制。第四、开放数据应基于“需求导向”，按照用户实际需求提供数据和进行数据主题分类，便于数据利用者发现和获取数据。更应随时关注数据的浏览量和下载量，动态调整开放数据的种类、数量和格式，基于数据利用者的实际需求而非政府部门的自我判断来开放数据。

随着我国政府数据开放工作的全面开展和深入推进，建议未来研究可持续跟踪我国开放数据的数量、质量与价值。由于目前我国已开放数据的地方政府数量有限，本研究还未能发现我国不同地区和层级间的显著差异，未来可在具备更多地方样本的基础上，比较我国不同地区和层级间的差异，发现影响我国政府数据开放和利用效果的因素及机理，探究推进政府数据开放的路径。

参考文献

- 夏义堃(2015). 国际组织开放政府数据评估方法的比较与分析. 图书情报工作, 19: 75-83.
- 郑磊(2015a). 开放政府数据的价值创造机理: 生态系统的视角. 电子政务, 7: 2-7.
- 郑磊(2015b). 开放政府数据研究: 概念辨析、关键因素及其互动关系. 中国行政管理, 11: 13-18.
- 郑磊、高丰(2015). 中国开放政府数据平台研究: 框架、现状与建议. 电子政务, 7: 8-16.
- 郑磊、关文雯(2016). 开放政府数据评估体系、指标与方法研究. 图书情报工作, 18: 43-55.
- Berners-Lee, T. (2014). *Linked Data*. Available at (Jan. 13, 2017): <https://www.w3.org/DesignIssues/LinkedData.html>.
- Open Government Working Group. (2007). *The Annotated 8 Principles of Open Government Data*. Available at (Jan. 13, 2017): <http://opengovdata.org/>.
- Open Data Charter. (2015). Available at (Jan. 13, 2017): <http://opendatacharter.net/principles/>.
- Viscusi, G., Spahiu, B., Maurino, A. & Batini, C. (2014). Compliance with Open Government Data Policies: An Empirical Assessment of Italian Local Public Administrations. *Information Polity*, 19(3 A): 263-275.
- World Bank. (2017). *Open Data Essentials*. Available at (Jan. 13, 2017): <http://opendatatoolkit.worldbank.org/en/essentials.html>.

to explore the role of government social media in Internet governance through combining big data analytics and small data analysis. To be specific , we focus on the content it spreads , the functions it performs , and the influence on Internet governance. Government social media disseminates various topics , of which propaganda of political values is the most notable and still increasing. The core function of government social media is servicing for people’s livelihood , while there exist the dual functions namely government affair openness and performance demonstration. Collecting the attention of 16% Chinese netizens , government social media is exerting its influence on kinds of fields. Government social media affects netizens’ political attitudes and behaviors by initiating their interest towards public affairs. However , its influence is bounded by local economic and social conditions , its communication strategies , and the netizens’ characteristics. The improvement of government social media should increase its coverage , optimize the communication strategy and improve the quality of information.

Key Words Information Dissemination; Network Governance; Government Social Media; Big Data; Small Data

Study on Open Local Government Data in China: Technical and Legal Characteristics

..... Lei Zheng & Jiuyang Xiong

Abstract This paper conducts a comparative study on the technical and legal characteristics of open data available on thirteen local government platforms in China in terms of the quantity , type , format , access method , timeliness , open license , metadata , click views and downloads. The study finds that the quantity of open datasets is pretty low in all regions and cities of China. Among those open government datasets , only 66. 85% are machine readable , and 18. 78% are in open formats. 66. 7% of them are committed to be updated regularly , however , only 20. 83% of them are actually updated. In many regions and cities , the right to access , use , and redistribute data freely is not explicitly guaranteed. The quality of metadata for open datasets varies from one region to another. In addition , the supply and demand for open data are not well matched. Finally , the paper puts forward some suggestions on the standard formats , open license , metadata provision and demand-oriented provision of open government data.

Key Words Open Data; Local Government; Assessment; China