

# 2018 年度 中国政府开放数据利用报告

华中师范大学信息管理学院  
湖北省数据治理与智能决策研究中心  
国家社会科学基金重点项目 (17ATQ006) 课题组  
2019 年 7 月

---

## 版权声明

本报告版权属于华中师范大学信息管理学院及湖北省数据治理与智能决策研究中心，受国家自然科学基金重点项目“基于全生命周期的政府开放数据整合利用机制与模式研究”（项目编号：17ATQ006）资助。凡转载或引用本报告的观点、数据等，请注明“来源：华中师范大学信息管理学院、湖北省数据治理与智能决策研究中心”。

---

# 前言

数据资源作为国家战略性资源，是支撑国家治理体系和治理能力现代化的重要基础。随着我国信息化发展水平的日益提高，政府、公众和企业对政府数据资源的采集、挖掘、组织、治理和应用也不断深化。政府开放数据不仅在于公开政府数据的数量，更在于它的利用<sup>1</sup>。能否将开放后的数据进行有效的利用，才是使数据发挥其价值的关键。因此，在依法依规的前提下，政府应最大限度地开放所掌握的数据资源，调动社会参与主体的积极性，充分利用数据资源，释放数据价值。

2018年我国政府开放数据工作重点逐步从建设转变为建设利用兼顾。1月，中央网信办、发展改革委、工业和信息化部联合印发《公共信息资源开放试点工作方案》，确定在北京、上海、浙江、福建、贵州开展公共信息资源开放试点。针对开放工作中平台缺乏统一、数据缺乏应用等问题，政府要求试点地区建立统一开放平台，鼓励通过政府专项资金扶持和数据应用竞赛等方式，支持社会力量利用开放数据开展创新创业，促进数据的社会化利用<sup>2</sup>。2015年国务院出台的《促进大数据发展行动纲要》要求在2018年底前建成国家政府数据统一开放平台，并开展政府和社会合作开发利用大数据试点，完善金融、税收、劳动就业、收入分配等领域国民经济相关数据的采集和利用机制，推进各级政府按照统一体系开展数据采集和综合利用。因此，对我国政府开放数据利用的情况进行分析，有

利于针对性地推进政府数据资源的开放利用工作。

近年来，开放政府数据在全国迅速推进。政府开放数据平台陆续增多，地级市及以上开放平台数量从2017年的19个新增至2018年的62个。2018年政府开放平台的开放数量显著增加，在建成国家政府数据统一开放平台的道路上更近一步，试点工作进展顺利，为提高政府数据使用效益打下了坚实的基础。但当前以开放为主要目标的格局缺乏形成对政府数据有效利用的目标导向，已开放的政府数据存在不好找、不好看、不好用等问题。在此背景下，对我国政府开放数据的利用现状进行分析和评估，通过典型样本研究，对照相关政策要求和公众需求发现问题，并提出对策，有助于推进全国范围内政府数据资源开放与利用。基于此，华中师范大学信息管理学院与湖北省数据治理与智能决策研究中心联合发布国内首个政府开放数据利用报告《2018年度中国政府开放数据利用报告》（以下简称报告）。

报告由华中师范大学信息管理学院与湖北省数据治理与智能决策研究中心基于多年科学研究积累和数据的监测分析制作完成，受到国家社会科学基金重点项目“基于全生命周期的政府开放数据整合利用机制与模式研究”（项目编号：17ATQ006）的资助。报告将政府开放数据的利用主体界定为政府、公众、企业，主要从利用行为、利用效益等方面对主体的利用情况做了具体分析，并对利用数据的典型案例深入剖析，以期增强对政府开放数据利用的认识，为政府数据的有效利用提出合理化建议。

1 马仁杰，金一鼎．价值实现视角下政府数据利用路径研究[J]．图书馆学研究，2018(13):18+39-44.

2 中央网信办、发展改革委、工业和信息化部联合开展公共信息资源开放试点工作[EB/OL]. [2019-05-23]. [http://www.cac.gov.cn/wxb\\_pdf/20180105006.pdf](http://www.cac.gov.cn/wxb_pdf/20180105006.pdf).

# 目 录

<b>1 绪论</b> .....	<b>1</b>
1.1 利用背景、意义与作用 .....	1
1.2 利用定义 .....	2
1.3 年度主要政策 .....	3
<b>2 平台主要内容</b> .....	<b>6</b>
2.1 平台建设 .....	6
2.2 数据集主题 .....	8
<b>3 利用行为分析</b> .....	<b>14</b>
3.1 利用行为概况 .....	14
3.2 浏览行为分析 .....	15
3.3 下载行为分析 .....	16
3.4 利用行为综合分析 .....	17
<b>4 利用效益分析</b> .....	<b>21</b>
4.1 利用效益综合评价 .....	21
4.2 数据开放度 .....	24
4.3 数据可持续性 .....	25
4.4 数据可利用性 .....	27
<b>5 利用案例</b> .....	<b>29</b>
5.1 政府层面 .....	29
5.2 公众层面 .....	30
5.3 企业层面 .....	31
<b>6 结论与对策</b> .....	<b>33</b>
6.1 总结 .....	33
6.2 存在问题 .....	34
6.3 对策建议 .....	35
<b>附录</b> .....	<b>39</b>
1.1 样本选取与数据采集 .....	39
1.2 工具选取与研究方法 .....	39

# 绪论

## 1 绪论

### 1.1 利用背景、意义与作用

大数据时代，数据资源在国家政治、经济、社会等方面的发展中扮演着至关重要的角色，其价值也随着数据量的不断扩大以及数据类型的不断丰富而日益凸显。政府开放数据在促进国家发展、社会进步上具有不可替代的重要作用。

#### 1.1.1 利用背景

政府开放数据运动在全球范围内迅速兴起，世界各国政府相继实施政府开放数据计划，旨在通过网络门户或平台共享开放政府数据。政府机构往往是交通、健康和环境等多个领域中领先的创造者和数据收集者。政府通过开放数据计划共享数据资源，并鼓励数据利用主体开发新产品和服务。对于我国而言，2018年是《促进大数据发展行动纲要》中要求的开展政府、社会合作开发利用大数据工作的试点年，是推进各级政府按照统一体系开展数据综合利用的启动年。

大数据作为人工智能发展的重要基础，云计算作为人工智能发展的重要推动力，机器学习作为人工智能发展的重要条件，三大支撑促进了以人工智能为代表的新一轮信息科技和产业变革。这一变革为推动政府开放数据利用提供了广阔的思路，在带来政府职能转变的同时，也使得公众需求发生了变化。政府数据是实现“宏观调控、经济监管、社会管理、公共服务、生态保护”五大核心职能的基础，是实现跨地区、跨层级、跨部门、跨系统的数据共享和流通的关键所

在<sup>1</sup>。

#### 1.1.2 利用意义

政府开放数据的充分利用和挖掘具有重要的社会意义，能够创造社会价值、推动社会创新，实现经济和社会效益。政府数据具有体量大和可靠性高等特点，蕴含极大的社会价值，需要通过应用程序接口、竞赛、智慧城市等各种方式对开放数据充分利用。上海连续四年举行 SODA 开放数据创新应用大赛，每年都有许多成功的开放数据应用产品产生。例如，参赛者融合政府开放不同的旅游相关的数据资源，研发设计的“数游长三角”全域旅游大数据平台，为政府提供全景监控、为景区提供实时检测、为公众提供长三角的旅游推荐，创造了极大的社会和经济价值。同时，政府开放数据的利用有利于实现社会的可持续发展。社会各项活动的开展都离不开政府数据的支持。公众的交通出行、医疗健康、劳动就业、教育文化、环境保护等都需要政府数据。智慧城市建设也需要开放政府数据，如阿里 ET 城市大脑依据城市交通体检、城市警情监控、城市交通微控、城市特种车辆、城市战略规划等场景需求，充分利用政府开放各类数据，解决杭州交通拥堵等问题。

此外，政府开放数据的利用在改变人们的生产方式、生活方式和思维方式上具有重大的实践意义。政府开放数据利用的过程实质上是政府数据资源再加工的过程，能够增加全社会数据资源的丰富程度和使用价值。社会对政府所开放的数据进行开发和利用，有

1 翟云. 重塑政府治理模式：以“互联网+政务服务”为中心[J]. 国家行政学院学报, 2018(06):128-132.

# 绪论

助于制定和改进开放政府数据政策，有利于将企业的生产方式从业务驱动转变为数据驱动，在此过程中，也改变了人们看待、接收和使用数据驱动的产品和服务的方式。

## 1.1.3 利用作用

### (1) 助力科学预测、实现精准决策

通过对政府数据的有效整合、共享开放和深化利用，提高政府数据的关联分析能力，有利于实现“用数据预测、用数据决策、用数据治理”的政府管理机制，实现基于数据的科学预测和决策，构建数据驱动的国家治理体系，推进国家治理能力现代化。在数据相对缺乏的时代，政府决策者在预测事物发展趋势时主要结合自身知识和经验基础，本质上是一种经验式决策<sup>1</sup>。通过开放的政府数据，政府可以基于全面准确的数据做出科学的分析，从经验式决策走向科学式决策，及时识别公众需求，从而更加精确地预测和决策。

### (2) 优化业务流程、提高服务效率

政府开放数据有助于优化企业业务流程，提高政府的服务效率。政府发布开放数据后，企业通过快速寻找并掌握开放的数据资源，利用已有的数据分析和处理能力，将数据转换为有利于事务决策的有效信息，从而优化企业的业务流程。此外，政府开放数据催生了一些企业通过提供第三方数据服务盈利<sup>2</sup>。这类企业

本身并不创造数据，而是通过搜集整合数据，搭建数据交易平台，以方便公众利用的方式满足其数据需求。例如，众多微信公众号及APP充分利用工商部门的企业信息为公众提供企业登记信息查询服务，通过地铁及公交线路信息的再利用提供查询、提醒、电子票进站等服务，在方便公众的同时提高了政府的服务效率。

### (3) 创新服务模式、推进数据赋能

政府开放数据利用通过政府和企业服务模式的创新，推进数据赋能。政府开放数据对于政府服务模式的创新，不仅体现为政府数据的全面整合、政务资源全面共享，行政组织结构、服务流程的优化，更为重要的在于政府服务理念的转变。政府通过对海量数据的深度分析与挖掘，主动识别公众的服务需求，打破政府各层级各部门之间的边界，提供无缝服务。以行政管理理念的转变和行政组织结构再造为基础，创新政府服务模式。企业根据自身运营需求，利用政府开放数据并整合自身掌握的其他数据，开发相关应用进行创新，以满足和提升企业业务需要；或根据欲解决的企业问题，将企业自身数据向社会开放，通过举办数据应用大赛，吸引公众或其他企业进行数据开发利用，提出解决方案。通过对企业全方位、深度、动态、持续的改造，使企业能够不断迭代、升级、转型、优化、创新，提升业务的数据化水平和数据的业务化水平，实现真正的数据赋能。

## 1.2 利用定义

政府开放数据是指由政府、政府委托和控制的实体产生的能被任何人自由地利用，再利用和再分配的

1 刘晓洋. 思维与技术:大数据支持下的政府流程再造[J]. 新疆师范大学学报(哲学社会科学版),2016,37(2): 118-125.

2 夏义堃. 企业开放数据再利用的困境与对策分析[J]. 电子政务, 2018(8):69-80.

数据。政府开放数据运动以 2009 年美国奥巴马政府签署《透明政府备忘录》为兴起标志，而后国内外政府和学术界对其高度重视。据 2016 年 4 月发布的“开放数据晴雨表”全球报告（Open Data Barometer）显示，全球已有 114 个国家加入了这一行列。我国以 2012 年“上海市政府数据服务网”的上线试运行作为开放数据的标志，截止 2018 年 12 月，全国共有 31 个省级政府建立了地方性政府开放数据平台（港澳台地区除外）。从政府部门高度重视政府开放数据以来，政府开放数据的利用也随之成为研究热点。政府开放数据与利用是政府部门针对在职能履行过程中所存储的各种数据面向公众、企业等开放，供其使用及研究。

本报告所提出的政府开放数据利用是指已开放的数据资源满足人们需求和利用的情况与程度，其本质是资源的有效配置和使用。对于政府开放数据利用的定义，还可以从四个角度去理解：从政府的视角来看，政府开放数据旨在促进政府的开放与透明，将具有公共利用价值的数据及时向社会公众开放。目的是方便公众使用、促进企业创新、改善公共服务；从利用者的视角来看，开放数据的主要利用主体是社会公众、企业、社会组织和政府机构，目的是实现数据的增值利用；从利用者的目标来看，社会公众对于政府开放数据的利用目标主要包括：①通过某个数字或者文字信息得到某种事实；②将数据转为可视化的图标用来提供信息；③通过新的界面相互获取数据并发现更多的数据集；④共享整合一些处理过的数据；⑤将数据隐藏在后台用来服务于某些场景；从利用的方式来看，政府开放数据倡导直接以数据集的形式通过互联网提供浏览和下载，也有其他方式如对公众开放

API（应用程序接口），提供查询接口，鼓励社会进行二次开发，还可以直接提供基于政府数据的应用软件。

### 1.3 年度主要政策

#### 1.3.1 国家政策取向

随着信息技术的高速发展，政府开放数据成为国家重要的基础性战略资源。我国也不断加强对政府开放数据发展的顶层设计，相继制定了一系列关于政府开放数据的政策，对促进我国政府开放数据的发展有着重要而深远的意义。2018 年度我国出台的国家层面的政府开放数据相关政策共计 9 条，为历年最高。政策内容目的主要包括提升政府网站管理水平和政务服务能力，建设整体联动、高效惠民的网上政府；进一步推动社会公益事业建设领域政府信息公开工作，推进国家治理体系和治理能力现代化；进一步加强和规范科学数据管理，保障科学数据安全，提高开放共享水平，更好支撑国家科技创新、经济社会发展和国家安全；做好政府公报工作，着力将政府公报打造成权威、规范、便民的政务公开平台；大力推进决策、执行、管理、服务、结果公开，不断提升政务公开的质量和实效，推动转变政府职能、深化简政放权、创新监管方式，促进经济社会持续健康发展，助力建设人民满意的服务型政府；全面贯彻习近平新时代中国特色社会主义思想 and 党的十九大精神，落实党中央、国务院关于推进依法行政、建设法治政府的部署和要求，切实保障群众合法权益，维护政府公信力；进一步深化“互联网+政务服务”，充分运用信息化手段解决企业和群众反映强烈的办事难、办事慢、办事繁的问题；

# 绪论

为深入推进“放管服”改革，全面提升政务服务规范化、便利化水平，更好为企业和群众提供全流程一体化在线服务，推动政府治理现代化；推动政务新媒体健康有序发展，政策详情见附录。本节通过文本分析的方式分析 2018 年度我国政府出台的政府开放数据相关政策，得到政策文本高频词分析结果，如表 1-1 所示。

报告发现“服务”一次共出现 444 次，频率为 2.0434；“平台”一词共出现 208 次，0.9572；“公开”一词共出现 183 次，频率为 0.8422；“管理”一词共出现 173 次，频率为 0.8808。结合以上 9 条政策的主要目的，可以看出我国 2018 年度对于政府开放数据的战略重点主要

为政府开放数据的平台建设与管理，更好的利用政府开放数据；同时力求全面高效的进行政府信息公开，其主要目的为建设人民满意的服务型政府，增强公民对政府开放数据的利用，维护政府的公信力。

## 1.3.2 地方政策取向

为响应国家政府开放数据战略，加快各地区政府开放数据发展，各地纷纷出台针对本地区的政府数据发展行动和规划，呈现出从中央总体规划逐渐向省级政府部门延伸的趋势。目前，已有 31 个省级政府出台

表 1-1 国家政府开放数据政策文本高频词 ( 出现频率 > 0.2500 )

序号	字词	出现频次	出现频率	序号	字词	出现频次	出现频率
1	服务	444	2.0434	14	资源	81	0.3728
2	政务	417	1.9191	15	共享	80	0.3682
3	平台	208	0.9572	16	办事	72	0.3314
4	数据	206	0.948	17	发展	71	0.3268
5	信息	196	0.902	18	媒体	71	0.3268
6	公开	183	0.8422	19	推动	71	0.3268
7	管理	174	0.8008	20	建立	69	0.3175
8	建设	173	0.7962	21	群众	69	0.3175
9	社会	137	0.6305	22	规范	66	0.3037
10	安全	103	0.474	23	系统	60	0.2761
11	科学	94	0.4326	24	网站	57	0.2623
12	推进	91	0.4188	25	公报	56	0.2577
13	经济	89	0.4096	26	实现	56	0.2577

了政府开放数据相关意见、方案或规划等政策文件，呈现出全面开花的格局，为政府开放数据发展营造良好的氛围。具体政策内容通过人工筛选结果见附录。我国各省份 2018 年度关于政府开放数据政策文件数量示意图，如图 1-1 所示。

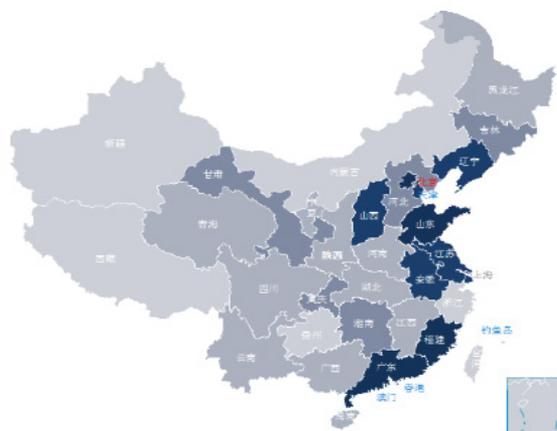


图 1-1 我国各省份 2018 年度关于政府开放数据政策文件数量示意图

为探究我国地方政府对于政府开放数据政策的发展趋势与取向，同时探究全国各地政府是否形成以国家政策为中心多点开花全面发展的局面。本节通过文本分析的方式分析 2018 年度我国地方政府出台的政府开放数据相关政策，得到政策文本高频词分析结果，如表 1-2 所示。

报告发现“公开”一词共出现 2073 次，频率为 1.3219；“服务”一词共出现 1944 次，频率为 1.2389；“互联网”和“网络”分别出现 814 和 795 次，频率

分别为 0.5173 和 0.5045。通过对比国家政策发现地方政府政策取向与国家政策取向有相似的趋势，表明地方政府开放数据政策紧紧围绕在国家政策下。同时，现阶段政府开放数据政策的重点偏向于网络建设，尤其是在大数据的环境下，利用网络来改善政府开放数据的利用情况。

表 1-2 地方政府开放数据政策文本高频词（频率 > 0.3000）

序号	字词	出现频次	出现频率	序号	字词	出现频次	出现频率
1	信息	2641	1.6859	15	建设	676	0.4279
2	公开	2073	1.3219	16	管理	652	0.4151
3	服务	1944	1.2389	17	推进	641	0.4087
4	政务	1163	0.7408	18	规定	608	0.3832
5	政府	1158	0.7344	19	经济	590	0.3768
6	安全	1012	0.645	20	发布	578	0.364
7	负责	1003	0.6386	21	提供	562	0.3576
8	落实	987	0.6258	22	债券	562	0.3576
9	监督	831	0.53	23	牵头	504	0.3193
10	公共	823	0.5237	24	网上	504	0.3193
11	互联网	814	0.5173	25	政策	504	0.3193
12	网络	795	0.5045	26	事项	489	0.3065
13	检查	789	0.4981	27	措施	471	0.3001
14	信用	782	0.4981	28	加强	471	0.3001

# 平台主要内容

## 2 平台主要内容

### 2.1 平台建设

政府开放数据平台是由政府牵头，各政务部门共同参与建设的平台，致力于各政府部门可公开数据的下载和服务，为企业和个人开展政务信息资源的社会化开发利用提供数据支持，推动信息资源增值服务业的发展以及相关数据分析与研究工作的开展。

基于我国政府开放数据平台大多数是以“data.gov.cn”为域名，因此，本报告以“data.gov.cn”为域名进行搜索，截止2018年12月，全国共有62个省、市或区政府建立了地方性政府开放数据平台（港澳台地区除外），详细情况见附件，2018年政府开放数据平台地理分布如图2-1所示。本节将从平台数量、类型、行政级别分布及地区分布四个方面分析2018年政府开放数据平台的建设情况。



图 2-1 2018 年政府开放数据平台地理分布图

从平台上线数量来看，2018年我国新增政府开放数据平台36个，平台上线数量同比增长138.46%，当前政府开放数据平台建设呈现出高速发展态势。平台数量的具体增长情况如图2-2所示。

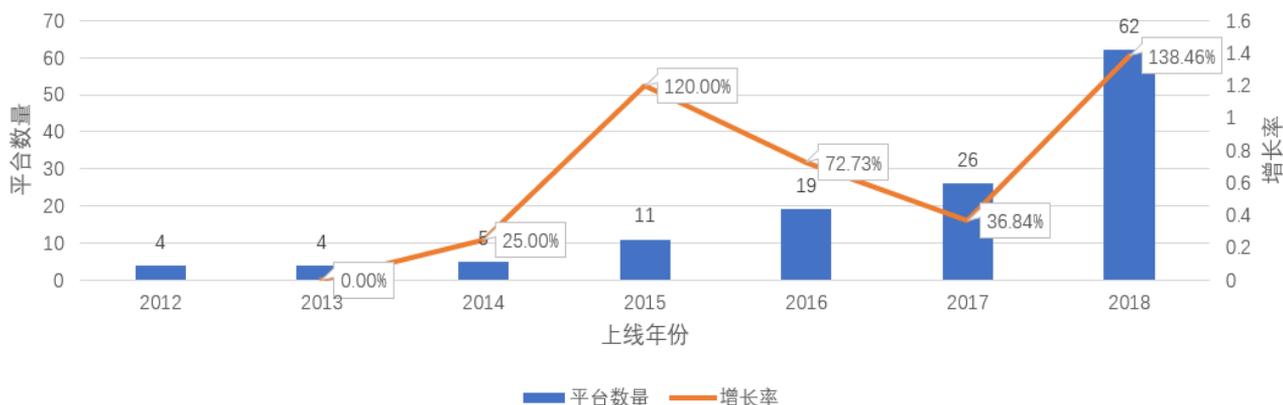


图 2-2 2012-2018 年国内开放数据平台上线数量与增长率统计图

从平台所属的地方政府行政层级来看，目前在已上线平台的各级地方政府中，市级数量最多达 44 个，省级行政区和副省级城市数量相持，分别为 8 个和 10 个。据统中华人民共和国行政区划统计，全国共有 31 个省级行政区（港澳台除外）、15 个副省级城市和 333 个地级行政区，其中除副省级城市平台接近半数外，我国总体上上线政府开放数据平台的地方占比依旧偏低。具体如图 2-3 所示。

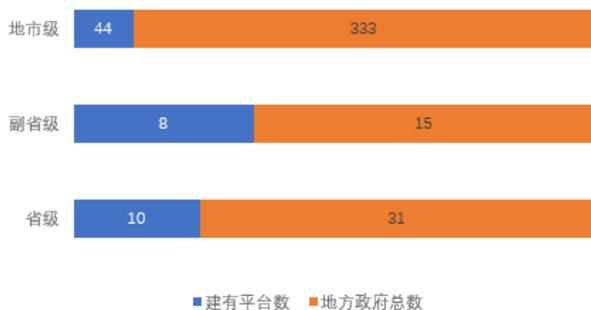


图 2-3 2018 年国内开放数据平台行政层级统计图

从平台类型来看，我国政府开放数据平台大多为专有式类型，即数据统一汇聚在一个专门的平台上进行开放，仅浙江、梅州、阳江、肇庆地区的政府开放数据平台类型为嵌入式，梅州市政府开放数据平台内嵌于梅州市人民政府网站的“政务公开”栏目，肇庆市政府开放数据平台内嵌于肇庆市人民政府网站的“用数据”版块，其专业性和便捷性都落后于专有式平台。需要特别说明的是，虽然荆门市政府开放数据平台标题为“荆门市人民政府”，但就其内容和结构看，完全具备专有式平台的特征，故将其划分到专有式平台中。2018 年国内开放数据平台类型统计详情如图 2-4 所示。



图 2-4 2018 年国内开放数据平台类型统计图

从平台所属地区来看，我国政府开放数据平台主要集中在分布于华东、华南地区，其中华东地区的平台数量高达 24 个，占国内平台总量的 52.17%，华南地区的平台数量为 12 个，占国内平台总量的 26.08%，华东华南地区的平台上线数在全国范围内处于领先地位。详情如图 2-5 所示。

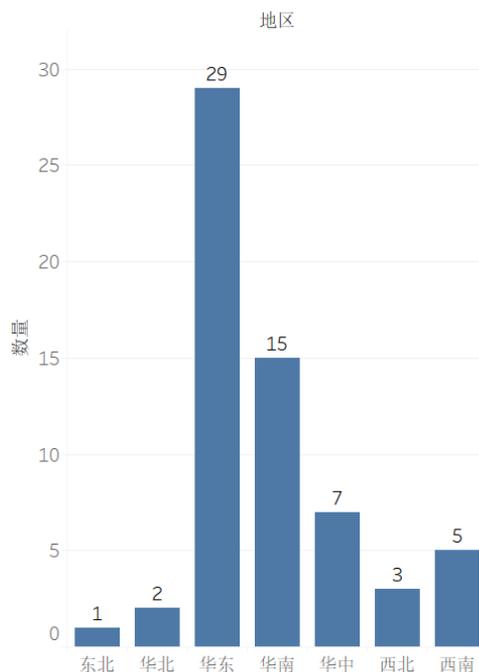


图 2-5 2018 年国内开放数据平台所属地区统计图

# 平台主要内容

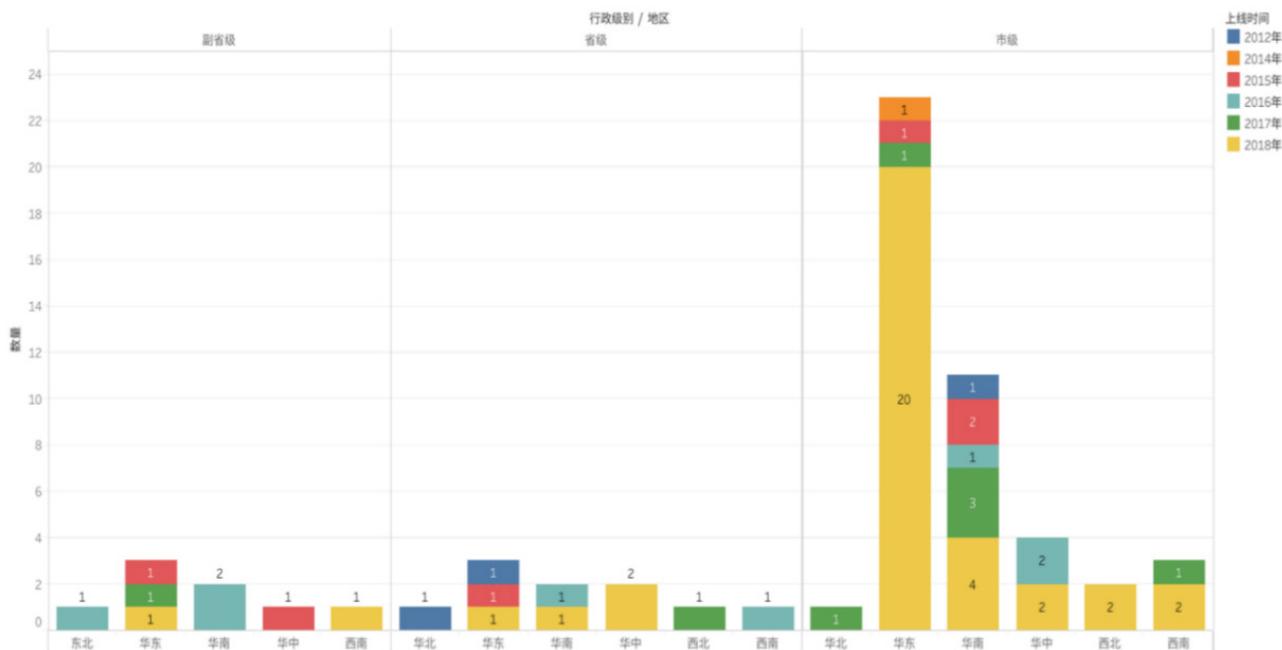


图 2-6 2012-2018 年国内开放数据平台所属地区与行政层级统计图

如图 2-6 所示，结合现有平台的所属区域和行政层级可以了解到，平台总量排名前三的地区分别是华东、华南、华中。其中，华东地区的政府开放数据平台数量为 29 个，在全国范围内处于领先地位，值得关注的是，2018 年该地区新增包括滨州、德州、东营等市级政府开放数据平台共 20 个，占 2018 年全年上线平台总量的 55.56%，市级平台建设速度较快，副省级、省级平台建设速度较缓；华南地区现有政府开放数据平台共计 15 个，市级平台建设较省级、副省级平台具备连贯性，但建设速度进展较为缓慢；华中地区政府开放数据平台上线总量虽排名第三位，但 2018 年仅襄阳上线了政府开放数据平台；其余地区现有的政府开放数据平台总量与 2018 年期间平台上线数均有待提升。

总的来说，2018 年华东地区和华南地区的政府开放数据平台建设情况较好，并呈现出将建设重点转移到市级平台的趋势，这一趋势在华东地区表现最为明显；华中、华北、东北、西南、西北地区平台建设情况较为缓慢。

## 2.2 数据集主题

我国政府开放数据的研究起步较晚，目前仅北京、广东、贵州等 62 个地方性政府建立了政府开放数据平台，由于行政级别存在差异，各开放平台开放数据的侧重点与进程差别较大，不适合做全样本分析。因此，经过多次讨论，本报告拟选取广东、上海、武汉、广州、哈尔滨、济南、贵阳和长沙 8 个城市为研究样本。笔者梳理了上述八个数据开放平台的主题分类情况，具



# 平台主要内容

## 2.3 数据集数量

从国内现有政府开放数据平台来看，数据集、数据接口和数据应用是平台开放的主要内容，可以直观

反映出平台开放数据的规模，2018年各平台的数据集情况如表2-2所示，其中数据集总量是数据集、数据接口与数据应用三者的数量之和。

表 2-2 2018 年政府开放数据平台数据集情况

平台名称	数据集	数据接口	数据应用	数据集总量
贵阳市政府开放数据平台	2913	480	15	3408
开放广东	2731	94	60	2885
武汉政府公开数据服务网	2203	14	54	2271
上海政府数据服务网	1364	646	89	2099
广州市政府数据统一开放平台	1237	778	8	2023
哈尔滨市政府开放数据平台	1062	2283	4	3349
长沙数据开放	320	236	6	562
济南市公共数据开放网	477	2897	18	3392

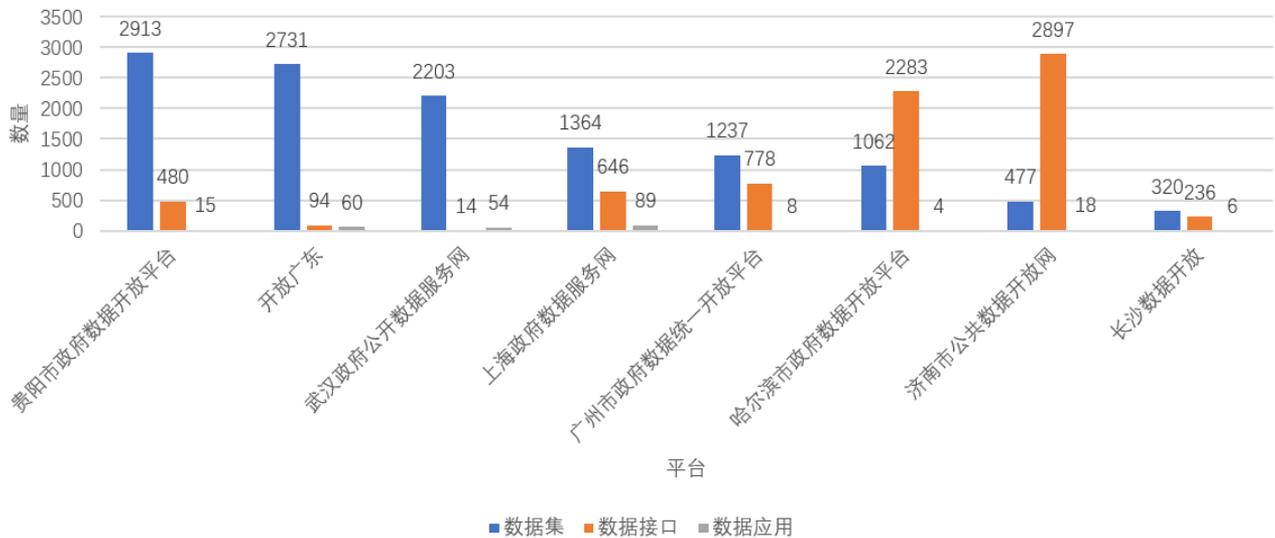


图 2-8 2018 年政府开放数据平台数据集数量统计图

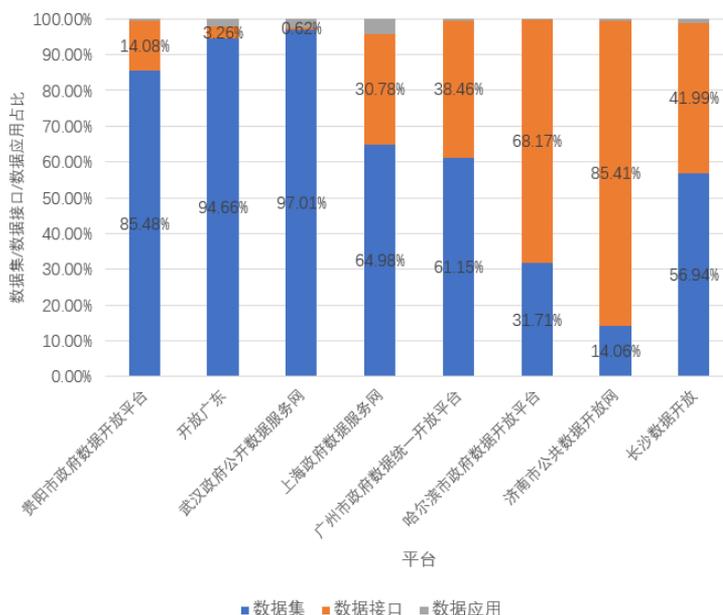


图 2-9 2018 年政府开放数据平台数据集 / 数据接口 / 数据应用占比统计图

从整体上看，各地平台开放的数据集数、接口数与应用数呈现不均衡分布，数据应用数量较少。目前贵阳、广东、武汉所开放的数据集数量排在前三位，开放数据集基数较大，但所开放的接口数占比极小；相比之下，济南和哈尔滨开放的数据集数量虽少，但所开放的接口数较多。此外，上述平台所开放的的应用数量远远低于数据集数量和接口数量，上海政府数据服务网的应用数最多（89 个），仅占 2018 年上海市开放数据集总量的 4.24%。具体如图 2-9 所示。

## 2.4 数据集格式

开放数据所采用的数据格式应当是开放的格式。开放格式是指文件不需要任何指定或付费的应用程序即可进行访问，由开放透明的过程定义数据格式标准并且不限制任何人实现其标准。此外，机器可读性作为开放需具备以下 3 项基本元素之一，要求数据若开放，则应是机器可读格式。常见的可机读格式与开放格式，如表 2-3 所示。

表 2-3 可机读格式与开放格式分类表

格式	开放格式分类
可机读格式	xlsx、xls、csv、kml、wms、wfs、txt、rft、xml、json
非可机读格式	pdf、doc、jpg
开放格式	json、csv、xml、kml、wms、wfs、txt、rtf
非开放格式	xls、xlsx、odt、doc、docx、pdf

# 平台主要内容

报告根据 Berners — Lee 提出的开放数据五星标准，选取广东、上海、武汉、广州、哈尔滨、济南、贵阳和长沙 8 个城市为研究样本，对各地区政府开放数据平台的数据格式进行梳理分级。其中，一星级数据指在互联网上开放授权以任意格式存在的数据，如 doc、pdf 等格式的数据；二星级数据指可机读的数据，如 xlsx、xls 格式的数据；三星级数据是指以非专属格式

开放的数据，如 csv 格式的数据；四星级数据指的是在三星级数据基础上利用 W3C 开放标准呈现的数据，如可以提供相应接口的数据；五星级数据则是在以上开放数据标准的基础上，提供数据发现服务。表 2-4 为各地区政府门户网站开放数据采用的格式以及分级情况。图 2-10 为基于 Berners — Lee 五星标准的 2018 年政府开放数据平台数据星级统计图。

表 2-4 2018 年政府开放数据平台数据类型统计表

平台名称	数据资源格式	一星级数据	二星级数据	三星级数据	四星级数据	五星级数据
开放广东	xlsx、csv、xml、json、api、xls、doc	41.26%	55.29%	0.04%	3.40%	0.00%
上海政府数据服务网	csv、docx、rar、xml、xls、xlsx、zip	5.40%	0.86%	8.71%	85.03%	0.00%
广州市政府数据统一开放平台	xls、xml、json、csv	1.01%	0.18%	98.80%	0.00%	0.00%
哈尔滨市数据开放	xls、xml、json、csv	7.23%	7.53%	85.24%	0.00%	0.00%
武汉政府公开数据服务网	xlsx、xls、jpg、png、docx、pdf、csv、mdb、swf	52.97%	13.75%	33.27%	0.00%	0.00%
济南市公共数据开放网	xls、xml、json、csv	1.68%	98.32%	0.00%	0.00%	0.00%
贵阳市政府开放数据平台	excel、csv、json、xml、rdf	0.00%	0.00%	100.00%	0.00%	0.00%
长沙数据开放	zip、xlsx	85.00%	15.00%	0.00%	0.00%	0.00%

由表 2-4 可知，各地区政府开放数据平台的数据类型呈现一定的共性与个性。共性体现在上述 8 个地区均无五星级数据，即在开放数据标准的基础上，未能提供有效的数据发现服务；个性在于各地区政府开放数据类型的分布呈现不均衡状态。从数据资源格式

看，广东、上海、武汉、贵州的开放数据格式较为丰富，能更好地满足不同需求的数据使用者更方便快捷地对数据进行增值利用；而长沙地区的政府开放数据格式仅有 zip 和 xlsx，数据类型较为单一。

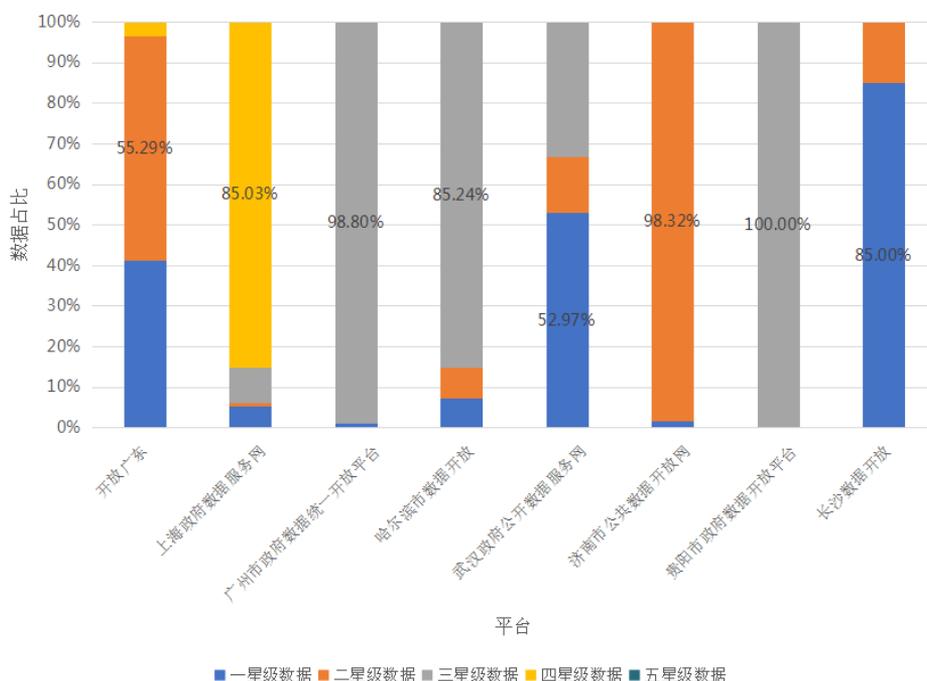


图 2-10 基于 Berners - Lee 五星标准的 2018 年政府开放数据平台数据星级统计图

由图 2-10 可知，上海市政府开放数据平台中的四星级数据占比最高，为 85.03%，上海市政府数据服务网的上线试运行作为我国开放数据的标志，其平台数据可机读性强、开放程度较高，在全国范围内处于领先状态；贵阳、广州与哈尔滨的开放数据平台中三星数据占比较高，分别为 100%、98.80%、85.24%，说明数据已具备较高的开放性，但利用 W3C 开放标准呈现的数据较少；广东、济南政府政府开放数据平台中二星数据占比较高，数据得可机读性较好，但开放性较差；武汉、长沙的政府开放数据中一星级数据占比最大，开放数据的可机读性与开放性都有待提升。

综上所述，2018 年是我国政府开放数据开放的关

键一年，根据 2015 年国务院印发的《促进大数据发展行动纲要》的要求，我国将于 2018 年底建成国家政府数据统一开放平台。在此背景下，本章对政府开放数据平台建设情况、平台主题分布、平台数据集情况以及数据类型情况进行了统计与可视化分析，得出目前我国政府开放数据平台层与数据层建设的相关结论如下：上线平台数量呈现高速增长态势，华东华南地区领跑全国，地区数据集数量规模化已有显现；开放数据主要集中在卫生、交通、教育、经济等社会生活领域的各方面，但各城市平台开放的数据主题在数量和名称上呈现差异；地区平台数据类型分布尚不均衡，数据发现服务暂且空白。

# 利用行为分析

## 3 利用行为分析<sup>1</sup>

### 3.1 利用行为概况

#### 3.1.1 主题视角下政府开放数据利用行为概况

为了揭示不同主题视角下开放数据的利用概况，首先统计各主题数据的浏览量、下载量，如图 3-1 所示；此外，分析单一及整体样本中开放数据的平均浏览率和平均下载率，如图 3-2 所示。由图 3-1 可知，主题视角下浏览量排在前三位的是经济工商、交通出行和民生服务，其中交通出行类数据的单一样本平均浏览率也是最高的，由此可知，经济工商和交通出行等主题的数据备受用户关注。此外，结合图 3-2 可知，仍然有部分主题在该指标的明显低于平均值，如政府机构与社会团体主题数据的平均浏览率均不足 300 次/条，这表明各主题数据被关注的程度存在差异性。

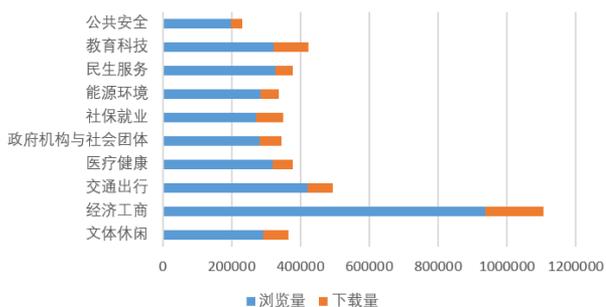


图 3-1 主题视角下开放数据的浏览量和下载量

1 段尧清, 邱雪婷, 何思奇. 主题与区域视角下我国城市政府开放数据利用现状分析[J]. 图书情报工作, 2018(20):65-76.

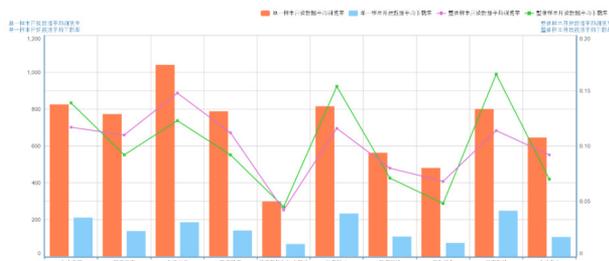


图 3-2 主题视角下单一（整体）样本开放数据的平均浏览（下载）率

#### 3.1.2 区域视角下政府开放数据利用行为概况

为了揭示不同区域视角下开放数据的利用概况，首先统计了各样本区域的浏览量和下载量，如图 3-3 所示；同时绘制了单一样本开放数据平均浏览率、下载率及整体样本开放数据平均浏览率、下载率统计图，如图 3-4 所示。由图 3-3 可知，区域视角下，浏览量排在前三位的城市是上海、贵阳、武汉；区域视角下，下载量排在前三位的城市是上海、贵阳、哈尔滨，由此可知，上海、贵阳等城市的政府开放数据备受用户关注。此外，结合图 3-4 可知，区域视角下，单一样本开放数据平均浏览率排在前三位的城市是上海、贵阳、广州，单一样本开放数据平均下载率排在前三位的城市是上海、贵阳、哈尔滨，这表明各区域数据被关注的程度存在差异性。

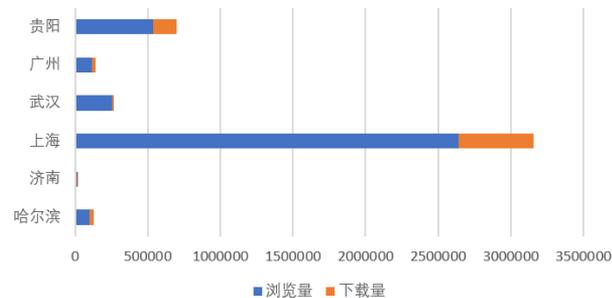


图 3-3 各地区整浏览量和下载量

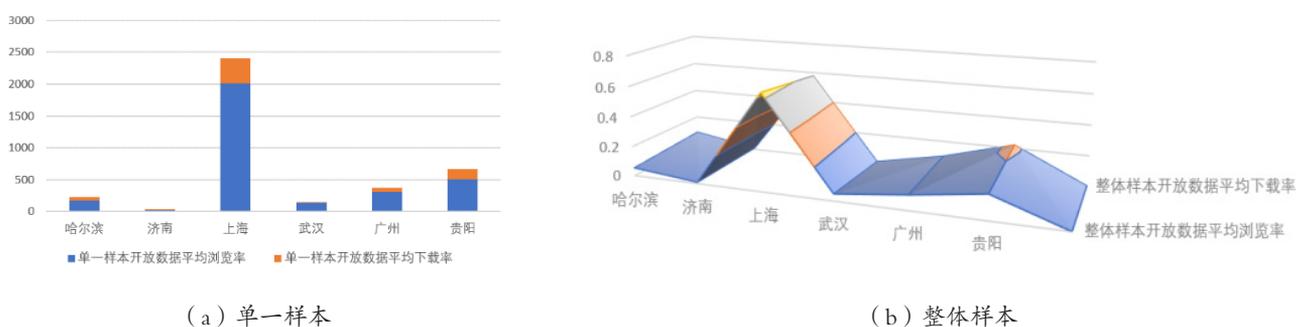


图 3-4 各地区单一（整体）样本开放数据平均浏览率和下载率

### 3.2 浏览行为分析

#### 3.2.1 主题视角下开放数据浏览率

浏览率能直观反映用户对某一主题数据的关注情况，首先运用 R 绘制各主题大类开放数据的浏览率折线图和浏览率散点图，如图 3-5 所示。其中 X 轴代表 10 个主题大类，也即  $t_i$  ( $i = 1, 2, \dots, 10$ )。主题 1-10 分别代表：文体休闲、经济工商、交通出行、医疗健康、政府机构与社会团体、社保就业、能源环境、民生服务、教育科技和公共安全。

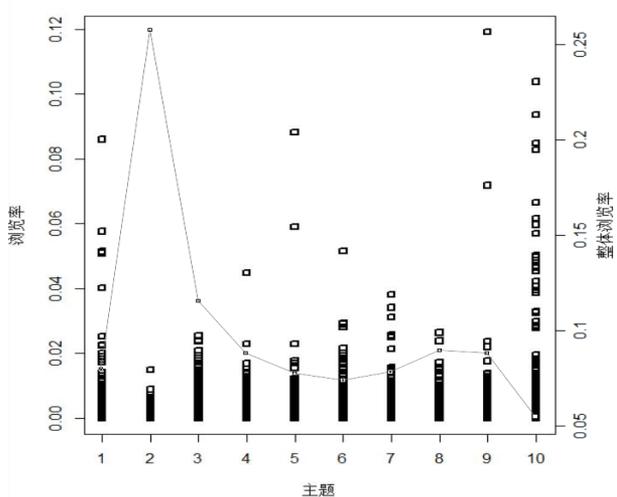


图 3-5 主题视角下开放数据浏览率折线图和散点图

图 3-5 显示浏览率最高主题 - 经济工商 (约 0.267)

是最低主题 - 公共安全 (约 0.054) 的 4.94 倍，且这 10 类数据中仅有经济工商和交通出行两类数据的浏览率高于平均值 (0.1)。此外结合图 3-1 和图 3-2，对比图 3-5 中的开放数据的浏览率折线图和散点图可知，经济工商类数据的浏览率虽然最高，但由其散点图的分布可知，并非其开放任意一条数据的浏览率都很高 (大多数落在 0-0.02 以内)。而公共安全类的数据却恰好相反，虽然该数据集的整体浏览率最低，但其开放数据浏览率的跨度较大，大多在 0.1-0.25 之间，这表明各主题开放数据整体与部分的特征并不具有一致性。

#### 3.2.2 区域视角下开放数据浏览率

同主题视角下开放数据的浏览率一样，将各区域的浏览率折线图与散点图绘制在一起，如图 3-6 所示，便于观察各地区开放数据的关注程度与状况。其中，X 轴代表 6 个城市，即  $c_i$  ( $i = 1, 2, \dots, 6$ )。区域 1-6 分别代表：哈尔滨、济南、上海、武汉、广州、贵阳。

# 利用行为分析

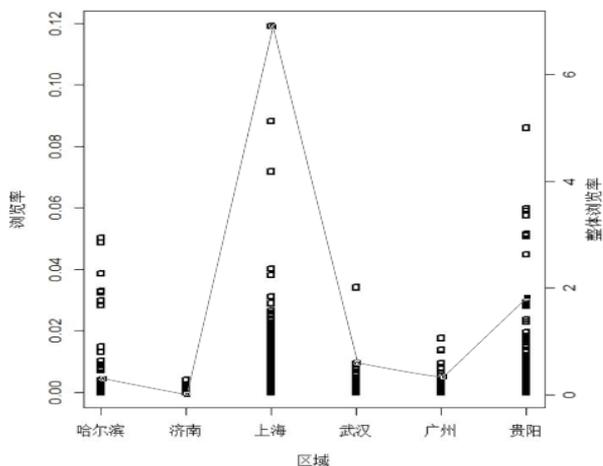


图 3-6 区域视角下开放数据浏览率折线图和散点图

结合图 3-4 和图 3-6 可知，不论是各区域浏览率还是其整体样本的平均浏览率，上海和贵阳都是表现最好的城市。以上海为例，其数据浏览率是济南的 166.88 倍，同时其整体样本的平均浏览率也达到 0.63，远高于平均值（0.1667）。而济南正好相反，这两项指标均排在末尾，其余各地排名稍有变化。此外，上海和贵阳两地开放的交通出行主题数据的浏览率之和占 6 地交通出行类数据总浏览率的 89.3%。

此外，结合图 3-6 的浏览率散点图可知，上海和贵阳不仅在整体上开放数据浏览率高，其散点图的跨度也较大。其中，浏览率落在平均值（1.67）以上的数据量较多；除了上海和贵阳以外，其余城市浏览率排名依次是武汉、广州、哈尔滨、济南，其中，哈尔滨虽然整体浏览率不高，但相应的散点图跨度较大，且其政府开放数据平台上存在相当数量的浏览率高于平均值的数据。

## 3.3 下载行为分析

### 3.3.1 主题视角下开放数据下载率

下载率是对浏览率的进一步说明与深化，它在很大程度上能反映出用户对某一数据的利用情况。同浏览率一样，图 3-7 所示的是各主题大类开放数据的下载率折线图和散点图。主题 1-10 分别代表：文体休闲、经济工商、交通出行、医疗健康、政府机构与社会团体、社保就业、能源环境、民生服务、教育科技和公共安全。

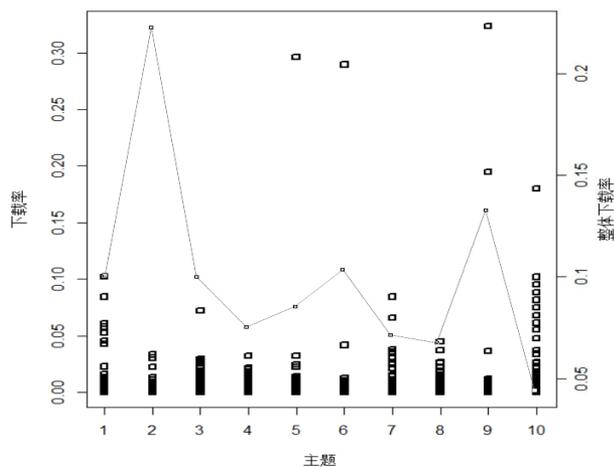


图 3-7 主题视角下开放数据下载率折线图和散点图

由图 3-7 可知，用户对经济工商、教育科技、社保就业和交通出行类数据的利用程度较深，体现在两个方面。第一，开放数据下载率高于平均值的有 4 个主题，即经济工商（约 0.223）、教育科技（约 0.133）、社保就业（约 0.103）和交通出行（约 0.1）；同时，除去社保就业外，其余 3 个主题的整体开放数据平均浏览率也均高于相应平均值，这间接表现出用户对此类数据利用的真实性。第二，经济工商等 4 个主题数据的下载率之和约为 0.56，占有主题开放数据下载

总量的一半以上，表明它们被利用的程度较高，同时也反映出其余几类数据的利用程度有待提高。

此外，结合图 3-7 中的下载率散点图可知，其余各主题下载率分布不均衡。首先，以公共安全为例，其下载率散点图的跨度虽然较大，但总体上该主题在下载率排在末尾；其次，包括经济工商在内的大部分主题整体开放数据的平均浏览率都不高，其值大多落在 0-0.05 以内。

### 3.3.2 区域视角下开放数据下载率

各区域开放数据的下载率反映了用户对各城市开放数据的利用程度。各区域下载率的折线图与散点图如图 3-8 所示。

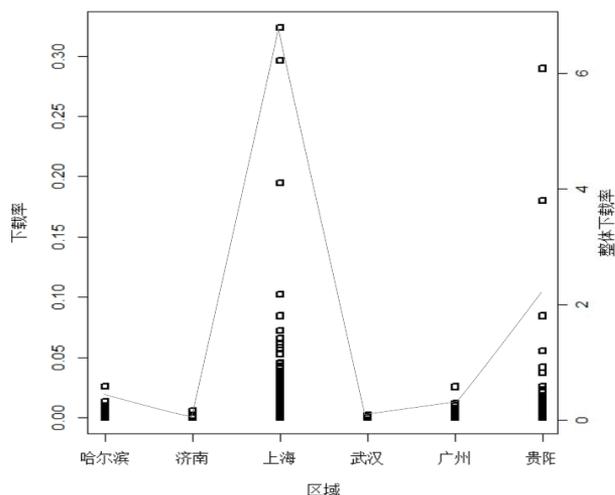


图 3-8 区域视角下开放数据浏览率折线图和散点图

由研究报告上文中图 3-4 可知，上海和贵阳分别作为一个整体，其下载率较高；同时其样本的平均下载率也占据了较大优势，例如上海作为开放数据下载率最高的城市，它与最低的济南相差 136 倍之多。此外，

上海除社保就业外的其他 9 个主题数据均拥有最高下载率，且全都超过平均值（约 0.167），其中教育科技的下载率更是高达 0.798。但同时也有城市的开放数据下载率较低，如武汉的开放数据在各项下载率的指标上均排名末尾。

结合图 3-8 中的各城市开放数据散点图可知，同各城市开放数据浏览率一样，下载率高的城市，其散点图的跨度也更大；反之，整体下载率低的的城市，其下载率跨度相对来说较小，但这并不代表各个城市开放数据浏览率与下载率的变化呈正比。

## 3.4 利用行为综合分析

### 3.4.1 主题视角下政府开放数据利用行为综合分析

为深入揭示各主题开放数据的利用现状与亲疏关系，对 10 个主题大类做聚类分析。聚类分析是指在事先不规定分组规则的情况下，将数据按其自身特征划分成不同的群组，各群组内部数据差距尽可能的小，而各群组数据之间的差距尽可能的大。首先选择聚类指标，由于单一样本和整体样本的开放数据平均浏览率（下载率）的变化方向一致且一一对应，前者是测算后者的基础，因此聚类分析主要参考开放数据浏览率、下载率和整体样本开放数据平均浏览率（下载率）4 个指标的相关情况。与此同时，选用层次聚类法，其中个体距离采用平方欧式距离，类间距离采用 Ward 联接，最终聚类结果如图 3-9 所示。

# 利用行为分析

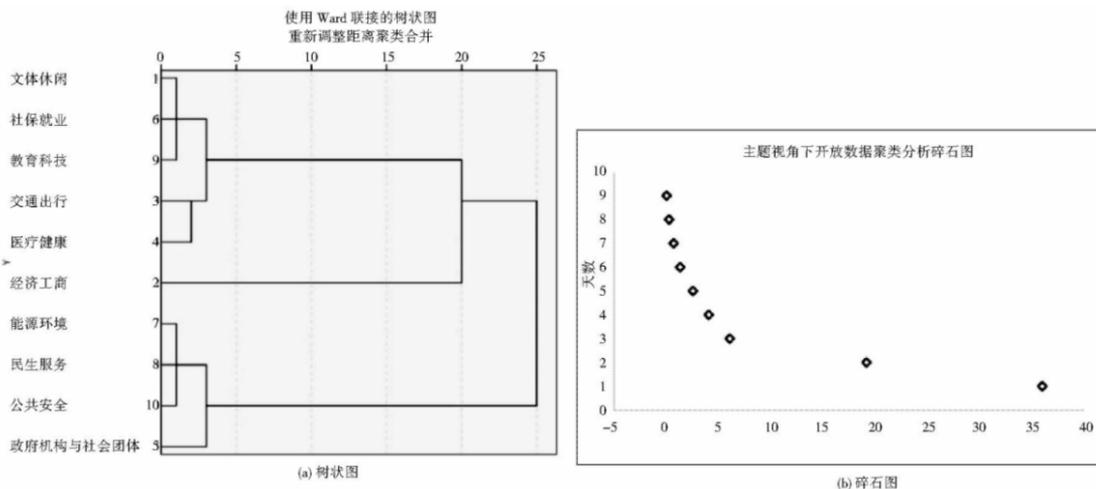


图 3-9 主题视角下开放数据聚类分析的树状图和碎石图

为了更好地划分聚类数目，绘制了主题视角下开放数据聚类的碎石图，如图 3-9 ( b ) 所示。随着类的不断凝聚和类目数量的不断减少，各类之间的距离迅速增大，碎石图逐渐趋于平坦。观察碎石图可知，当聚成 4 类之前，各类之间的距离较小；当聚成 4 类之后，各类之间的距离较大。由此可知，4 类就是该碎石图的“拐点”，因此聚成 4 类或 3 类较好。经过综合考虑，本报告最终将 10 大主题聚为 4 类，见图 3-9 ( a )。

第一类是“文体休闲”“社保就业”和“教育科技”。仅经过三步就聚成一类，其系数分别为 0.128 和 1.465。社保就业和教育科技类的数据在整体样本开放数据平均浏览率等 3 项指标均超过相应平均值，其余各项指标均排名较前，方差与标准差都较小。这表明与民生相关的数据利用效率较高，教育科技、社保就业和文体休闲是用户最为关心的日常问题之一，与用户的距离最近，也进一步说明用户需求是数据利用的前提。

第二类是“经济工商”。经济工商在浏览率、下载率等 4 项指标中占据绝对优势，因此自成一类。该类数据较高的关注度与利用率主要是由上海市政府的

开放数据贡献的，此类数据主要涉及某地经济建设和工商贸易等信息，涵盖了经济、工商、统计、贸易、消费、经济政策信息等方面。用户尤其是企业用户对此类数据的关注和需求更大，因此经济工商类数据各项指标都稳居高位。

第三类是“交通出行”和“医疗健康”类数据。它们在第 5 步时与医疗健康聚成一类，二者之间的系数为 5.187。这类数据虽然在总体上浏览率与下载率不高，但其整体样本开放数据平均浏览率与下载率却排在前列。交通出行与用户的生活联系紧密，而医疗健康更是全社会关注的热点，当下“互联网+”交通和电子医疗的出现，大大节约了用户的时间，便利了公众的生活。

第四类是“民生服务”“能源环境”“政府机构与社会团体”和“公共安全”类数据。它们的 4 项指标均为负，且低于平均值。这一类数据的利用相对不高，用户对能源环境、公共安全等社会治理领域的问题目前关注还不太多，与用户意识、需求的紧急性等因素有关。

### 3.4.2 区域视角下政府开放数据利用行为综合分析

为进一步揭示政府开放数据在区域视角下的利用现状，结合浏览率等 4 个指标，采用层次聚类法对各区域开放数据的利用状况做聚类分析。其中，个体距离采用平方欧式距离，类间距离采用平均组间联接，最终聚类结果的冰柱图和树状图如图 3-10 所示。

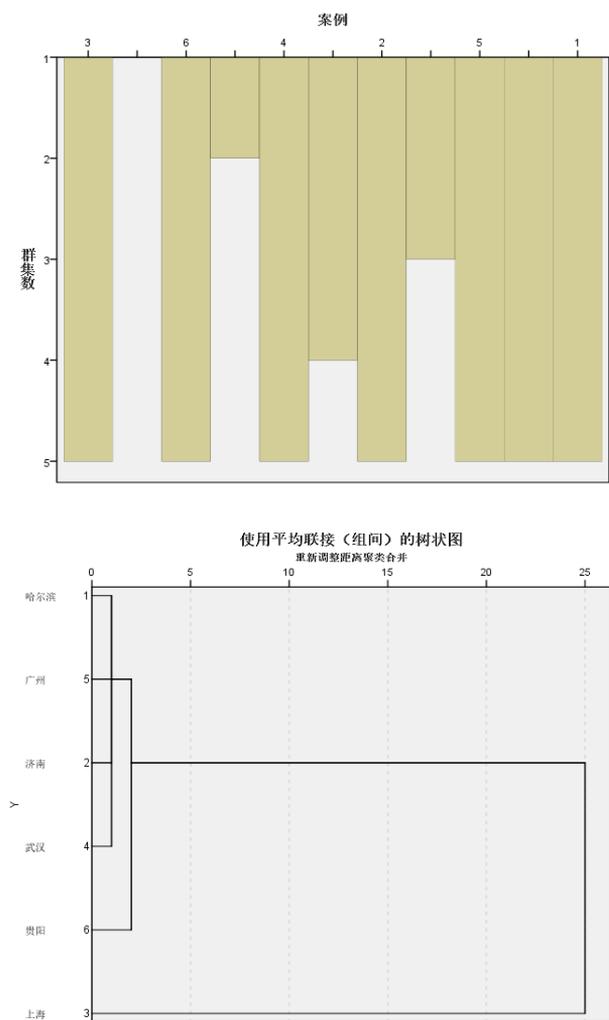


图 3-10 区域视角下开放数据聚类分析冰柱图和树状图

由于区域研究样本较少，因此直接采用观察法对

其聚类数目进行划分。据冰柱图可知，当聚成 4 类时，哈尔滨和广州为一类、济南和武汉为一类、贵阳和上海分别单独聚成一类；当聚成 3 类时，哈尔滨、广州、济南和武汉为一类、贵阳和上海分别单独为一类。为了更为细致的分析开放数据利用的分布研究，将 6 个研究样本聚成 4 类。

第一类是哈尔滨和广州。两地具有较高的相似性，首先聚为一类，系数仅为 0.043。在 4 个衡量指标上，哈尔滨和广州的指标排名相差不大，紧跟在上海和贵阳之后，各项指标大多处于中间位置。同时，哈尔滨和广州开放数据的相似性还体现在其数据开放的起始时间、开放数据的数量、格式等方面，虽然起步较晚，但其政府数据的关注度和被利用情况尚可。

第二类是济南和武汉。这两个城市之间的系数为 0.125，相似度较高。武汉虽然比济南早 2 年开放数据，开放数据集总量大，有较高的浏览率，但由于受到开放格式等因素的影响，武汉开放数据的整体和平均下载率都很低。同样，济南开放数据起步晚，数据集数量少，在 4 项指标中的排名都十分靠后。因此与武汉的差距并不明显，这两地聚为一类。

第三类是贵阳。贵阳虽然开放数据起步较晚，但在不足一年的时间内就取得数据开放指数排名第二的成绩。除整体样本的平均浏览率外，贵阳开放数据的其余各项指标均超过平均值，此外，其开放的 9 个主题数据集（民生服务主题暂缺）的浏览量均稳居前三，贵阳开放数据的关注程度和利用程度仅次于上海。

第四类是上海。上海开放数据的浏览率和下载率

# 利用行为分析

均在首位，自 2012 年开放数据以来，上海通过政府引导、提高数据质量、重视用户参与和数据创新等方式，使得其在各地方政府开放数据中稳居前列。本报告中，上海的开放数据在浏览率、下载率等 4 个指标中均排名第一，其中经济工商类数据的单一平均下载率高达 352 次 / 条，是其他类数据的 2 倍之多。

通过研究分析，本报告得出以下结论：第一，从主题视角看，用户对经济、民生等与日常生活联系密切的领域关注更高，10 个不同主题数据的利用现状呈

现出一定的差异性，将 10 个大类的数据根据相似性划分成了 4 类：文体休闲、社保就业和教育科技是第一类；经济工商单独成第二类；交通出行和医疗健康是第三类；民生服务、能源环境、政府机构与社会团体和公共安全聚成第四类。第二，从区域视角看，不同城市开放数据的利用程度不同，呈现出明显的高低之分：哈尔滨和广州为第一类、济南和武汉是第二类、贵阳和上海分别单独聚成第三类和第四类。

# 利用效益分析

## 4 利用效益分析

### 4.1 利用效益综合评价

政府开放数据利用效益综合评价主要是从微观层面的数据与用户视角出发，采用数据集个数、可机读格式、开放格式、数据接口 API、动态更新、静态更新、

APP 应用、评分量、浏览量、下载量这 10 个指标来衡量，各指标详细情况见附录。利用网络爬虫软件及人工观察的方式获取以上指标的相关数据，并采用熵值法计算各指标数据，得到政府开放数据利用效益评价指标体系的权重如表 4-1 所示。

表 4-1 政府开放数据利用效益评价指标体系权重

目标	一级指标	权重	重要性排序	二级指标	权重	重要性排序
政府 开放 数据 利用 效益	数据数量	0.101682654	4	数据集个数	0.101682654	1
	数据开放度	0.235186547	2	可机读格式	0.078167826	7
				开放格式	0.065772145	9
				数据接口 API	0.091246575	6
	数据可持续性	0.125916012	3	动态更新	0.057413245	10
				静态更新	0.068502767	8
	数据可利用性	0.537214787	1	APP 应用	0.122052964	5
				评分量	0.151104935	2
				浏览量	0.128377693	4
				下载量	0.135679195	3

表 4-1 给出了政府开放数据利用效益评价的各项指标及其权值，其中一级指标排序为数据可利用性、数据开放度、数据可持续性和数据数量。在这四项一级指标中，数据可利用性权值最大，说明政府开放数据利用效益评价中最重要的一级指标为数据可利用性，也说明了其评价与数据可利用性具有较高的关联性。数据可利用性的二级指标分别为 APP 应用、评分量、

浏览量及下载量，这些指标具有较高的权重，从一定程度上反映了用户对政府开放数据的需求。

根据表 4-1，可以计算得到各地政府开放数据平台利用效益得分及排名，如表 4-2 所示。结果表明，上海市政府开放数据利用效益得分最高，依次为广东省、广州市、贵阳市、哈尔滨市。

# 利用效益分析

表 4-2 各地政府开放数据平台利用效益得分及排名

省 / 市	排名	数据数量	数据开放度	数据可持续性	数据可利用性	利用效益得分
上海市	1	0.0183991	0.0398134	0.1250409	0.4838748	0.6671282
广东省	2	0.1016827	0.0781678	0.0525614	0.1848662	0.4172781
广州市	3	0.0106618	0.0864416	0.0951034	0.1645685	0.3567753
贵阳市	4	0.0769474	0.1451793	0.0346837	0.0593236	0.3161340
哈尔滨市	5	0.0000000	0.1358784	0.0685028	0.0092413	0.2136225

通过对政府开放数据利用效益的结果分析，可以得到以下四个结论。

(1) 政府开放数据利用效益与数据可利用性的关联性最高。数据可利用性所占权重最大，说明政府开放数据利用效益评价中最重要的一级指标为数据可利用性，而数据可利用性的二级指标为 APP 应用、评分量、浏览量、下载量，这些二级指标从一定程度上反映了用户对政府开放后数据的利用情况。从数据可利用性中可以看出，上海市的数据可利用性最高，哈尔滨市最差，这与数据的浏览量、下载量存在一定的关系。如相对于其余四个政府开放数据平台来看，上海市的浏览、下载情况最佳，而哈尔滨市政府开放数据平台数据集的浏览、下载情况较差，在一定程度上影响了数据可利用性的权重，也表明了哈尔滨市所开放的数据集主题、内容等不能满足用户的需求。

(2) 上海市的政府开放数据利用效益最佳。根据利用效益得分及排名可知，上海市政府开放数据利用效益最佳，依次为广东省、广州市、贵阳市和哈尔滨市。上海市政府开放数据平台的利用效益排名主要得益于平台的数据可利用性。在数据可利用性方面，上海市政府开放数据平台具有 APP 应用多、数据评分高、

数据易于下载和使用、浏览量和下载量高等优势，使其在数据可利用性方面具有最高的得分。而哈尔滨市政府开放数据平台由于其所提供的数据集个数、数据应用数及各数据集的浏览、下载情况都比较少，使其在数据可利用性、数据数量中所获得的权重比较低，让其在政府开放数据利用效益评价中处于劣势地位。

(3) 各地政府开放数据利用效益具有明显的层次性。根据政府开放数据利用效益得分，将这 5 个政府开放数据平台划分为三个层次：上海市政府开放数据平台遥遥领先于其余政府开放数据平台，划分为第一层次，广东省、广州市和贵阳市为第二层次，哈尔滨市单独为另一个层次。

(4) 政府开放数据利用效益的评价研究是一个动态变化的过程。这与数据提供者、门户网站以及用户的利用状况、关注度等都有一定的关系。随着大数据环境、政府机构的观念、用户需求的改变等，政府部门开放数据的方式、内容等都应随之不断变化，则评价指标体系和权重也应该在新的环境下进行调整和完善。

针对以上政府开放数据利用效益的评价指标数据及评价结果，对上海市、广东省、广州市、贵阳市和

哈尔滨市这五个政府开放数据平台提出些许建议。

(1) 对于上海市的开放政府数据实践来说,上海市是中国开放政府数据项目启动最早的城市之一,可以说是我国开放政府数据的最佳实践,其先进的做法,为其他地方的政府数据门户建设提供借鉴意义。上海市政府数据服务网在 2012 年 12 月底正式试运行,2015 年 5 月 7 日,上海市政府数据服务网 2.0 版正式开通,上海市政府数据资源网的开放内容现已基本覆盖各部门主要业务范围,涵盖资源环境、文化休闲、卫生健康、社会发展、民生服务、经济建设、教育科技、机构团体、公共安全、道路交通这 10 个主题大类,累计开放的数据集数为 1364 个。但其也存在不足之处,与其余四个政府开放数据平台相比,上海市开放的数据集个数相对较少,上海市应要扩大数据集规模;上海市开放的数据格式均为 XLS 和 CSV,在后期应要增加 XML、JSON、RDF 等格式的使用;上海市的数据集动态更新频率虽然比其他四个平台高,但是整体的更新频率比较低,应要加快数据集的更新频率,以便使用户能够快速、高效的使用这些数据集来创造更多的应用价值。

(2) 对于广东省的开放政府数据实践来说,广东省政府开放数据平台在数据开放度和数据可持续性存在不足。数据开放度的衡量指标为可机读格式、开放格式和数据接口 API,数据可持续的衡量指标为数据的更新频率,分为动态更新和静态更新。报告中,广东省所提供数据集的数据格式均为 XLS、XLSX 格式,均为 Excel 表格形式,均为机器可读格式,但并不是开放格式,该平台应将提供的数据集格式改为开放格式

CSV,而不是大部分均为 XLS、XLSX 格式;该平台开放数据集的更新频率多数为每年更新或不定期更新,不利于用户及时使用数据,应加快数据集的更新频率,如每季度、每月、每周或实时更新等;与其他四个平台相比较,广东省开放的数据集个数最多,但相应的数据接口数却最少,该平台应该多开发相应的应用程序编程接口,增加 API 数量,以及加强开发 APP 应用的能力。

(3) 对于广州市的开放政府数据实践来说,广州市政府开放数据平台在开放数据集数量、数据开放度、数据可持续性和数据可利用性均存在不足。广州市要逐步扩大其开放数据集的规模,特别是要扩大核心数据集的开放,在扩大数据集规模的同时,要增加数据可机读性、开放性以及数据接口数 API,同时需要加快数据集的更新频率等等。

(4) 对于贵阳市的开放政府数据实践来说,贵阳市政府开放数据平台在数据可持续性存在一定的不足。贵阳市最重要的是要提高数据可持续性,加快数据集的更新频率。报告中,贵阳市的动态更新与静态更新的比例为 0.049661,大部分数据都是每半年更新、每年更新或者不更新,只有极少部分为每季度更新、每月更新和实时更新,不利于用户及时使用数据。贵阳市应要加快数据集的更新频率,以使用户能够快速、高效的使用这些数据集来创造更多的应用价值。

(5) 对于哈尔滨市的开放政府数据实践来说,哈尔滨市政府开放数据平台在数据数量和数据可利用性存在不足。在数据数量方面,哈尔滨市开放的数据集

# 利用效益分析

个数最少，为 1062 个，应要扩大数据集的规模；在数据可利用性方面，哈尔滨市的浏览、下载情况最差，其提供的 APP 应用数也最少，表明该平台开放的数据集受用户关注程度较低，应从数据使用和下载的易用性方面入手，提高用户对数据集的满意程度。

## 4.2 数据开放度

数据开放度是指数据的开放性，主要用来衡量数据的开放程度，其二级指标包括可机读格式，开放格式以及数据接口 API。

### 4.2.1 可机读格式

数据开放应该是机器可读的格式，是用来衡量数据是否方便用户获取和利用、是否能被计算机自动读取和处理，如 XLS、CSV、TXT、XML、JSON、XLSX、WMS、WFS、RFT、XML、RDF 等格式，而 DOC、PDF、JPG 为机器不可识别读取的格式。因为 XLS、XLSX 为常用的 Office 软件格式，CSV 为自动化处理的结构化数据，故大多数政府开放数据平台一般提供两种及以上的数据格式。

### 4.2.2 开放格式

开放格式是指可下载数据集应以开放的、非专属的格式提供，任何实体不得在格式上排除他人使用数据的权利，以确保数据无需通过某个特定（特别是收费的）软件或应用程序才能访问。如 JSON、CSV、XML、TXT、KML、WMS、WFS、RTF 等，而 XLS、XLSX、DOC、DOCX、PDF、ODT 不属于开放格式。

### 4.2.3 数据接口 API

API 接口指平台向公众开放数据的 API 应用程序接口的情况，是否支持以数据调用接口的方式提供数据，满足部分用户开发应用程序的需求。

上海市、广东省、广州市、贵阳市、哈尔滨市这五个政府开放数据平台提供下载的数据集格式种类比较少，主要为 CSV、XLS、XML、JSON、XLSX、RDF 等数据格式，如表 4-3 所示。相对于美国政府开放数据平台上所提供的 40 余种数据格式和每一类数据集的多种格式来说，我国的政府开放数据平台所提供的的数据格式较为单一，且每一种数据集所选择的供应格式也比较单一。在上海市、广东省、广州市和哈尔滨市这四个政府开放数据平台建设中，均未能达到此水平，只有贵阳市提供的数据格式尚且多一些，贵阳市政府开放数据平台在数据开放度建设方面较为可行。

表 4-3 各地政府开放数据平台数据格式统计

平台	可机读格式	开放格式	格式
贵阳市政府开放数据平台	√	√	CSV、XLS、XML、JSON、RDF
哈尔滨市政府开放数据平台	√	√	CSV、XLS、XML、JSON
上海市政府数据服务网	√	√	CSV、XLS
广州市政府数据统一开放平台	√	√	CSV、XLS、XML、JSON
开放广东	√	√	XLS、XLSX

基于 Office 软件的 XLS、XLSX 等格式应用较为广泛，存在使用习惯和兼容性考量，此类数据集才会提供两种及以上类型的格式，在数据发布标准统一制定

后，这类数据集格式的问题可以被标准化。在贵阳市还广泛提供了 RDF 格式，RDF 是为数据集设置固定的 URL 链接，有助于用户发现和链接到目标数据集的具体位置。在贵阳、哈尔滨、上海、广东、广州这五个平台中，根据所提供的数据格式逐一进行判断是否符合

机读条件和开放条件，最后得到贵阳、哈尔滨、上海、广州、广东这五个政府开放数据平台上的数据可机读比例约为 100%，96%，44%，99% 和 100%；开放格式约为 100%，96%，44%，99% 和 0%，如图 4-1 所示。

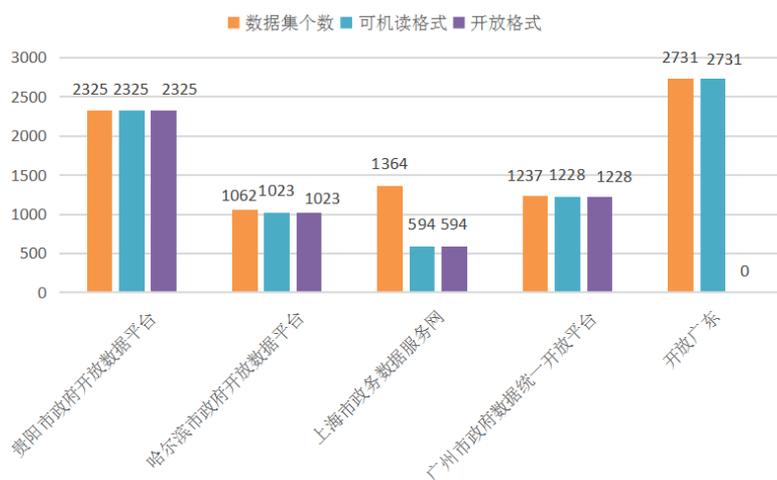


图 4-1 各地政府开放数据平台数据集的可机读格式和开放格式数量统计

从数据集的开放格式来看，广东省的开放格式为 0，这是由于其平台上提供利用的数据格式多数为 XLS、XLSX 格式。要想提高其开放性，就得将数据更多得以机器可读的数据格式提供利用，如 CSV、JSON、XML 等格式，结合广东省的数据接口数和数据应用来看，该平台还可以多开发相应的应用程序编程接口，以及加强开发 APP 应用的能力。

### 4.3 数据可持续性

数据可持续性用来衡量各平台数据集自创建以来是否进行数据更新，主要以数据的更新类型来衡量，更新及时性是由开放数据门户网站上的每一条数据更

新的及时性所决定的。本次主要将其划分为动态更新和静态更新。

#### 4.3.1 动态更新

评估政府开放数据平台上数据集是否“活跃”的标准之一就是数据的更新频率，动态更新的数据有利于反应事物的真实现状，从中挖掘出数据的价值。报告中将每季度更新、每月更新、每周更新、每日更新、实时更新等视为动态更新。

#### 4.3.2 静态更新

静态更新的数据不利于揭示事物的发展情况，

# 利用效益分析

更不利于数据价值的挖掘。部分政府开放数据平台所开放的数据集更新不及时或长时间不更新容易导致门户网站失去活力。从用户角度来看，若长时间在网站上找不到满足自身需求的数据，便会逐渐失去对政府的信心。报告中将不更新/一次性更新、不定期更新、每年更新、每十年更新、每五年更新、每半年更新、按需更新、自定义更新等视为静态更新。

贵阳市、上海市等五个政府开放数据平台上数据

集更新类型主要包括每半年更新、每年更新、每季度更新、每月更新、实时更新、不定期更新、一次性更新/不更新、每周更新、每天更新、自定义更新等其他多种更新类型。通过对各地政府开放数据平台数据集进行统计分析发现，仅有 9.11% 的数据可以每季度、每月、每周、每日、实时更新，其余 90.82% 的数据更新频率为每半年、每年、不定期、自定义或一次性更新，如表 4-4 所示。

表 4-4 各地政府开放数据平台数据更新频率分布

城市	每季度	每月	每周	每日	实时	每年	每半年	不定期	一次性	其他	自定义
上海市	102	181	0	0	1	938	81	0	0	61	0
广东省	47	207	0	0	6	1005	0	1466	0	0	0
贵阳市	62	44	0	0	4	1968	144	87	16	0	0
广州市	6	45	27	44	18	288	10	540	0	258	1
哈尔滨市	0	0	0	0	0	0	0	1059	0	3	0
合计	9.11%					90.82%					

同时，将各地政府开放数据平台上开放的数据集的更新频率进行汇总，统计结果如图 4-2 所示。从图 4-2 中可以发现，数据集更新类型大多数为每年更新和不定期更新，占比分别为 48.16% 和 36.15%。由此可见，当前在各地政府开放数据平台上已开放的数据多数为静态数据，数据更新不及时，严重限制了数据利用主题对各数据集的开发利用。

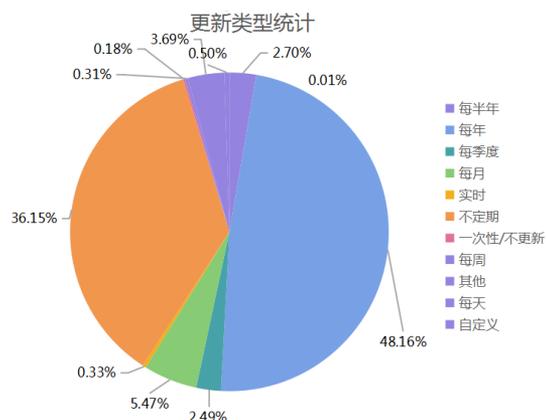


图 4-2 各地政府开放数据平台数据集更新频率统计

#### 4.4 数据可利用性

APP 的应用、数据的评分量、数据的浏览量和下载量是体现用户对数据的关注度和数据价值的重要表现形式。对于数据可利用性的评估，主要是从微观层面的用户视角来进行分析，当前研究大多数都从浏览量、下载量、评分量等角度来衡量开放数据的利用效果，故数据可利用性主要是以 APP 应用、评分量、浏览量及下载量来衡量。

##### 4.4.1 APP 应用

应用开发情况指的是将政府开放数据转化为实际应用的情况，只有将这些政府数据转化为便于实际用户使用的应用，或者对数据进行进一步的挖掘分析出有用的价值，政府开放运动的意义才得以体现。APP 应用是指各平台是否有将数据开发成相应 APP 应用程序，为公众提供便民服务的功能。报告中的 APP 应用评价主要是以各政府开放数据平台已经公布的 APP 应用个数来进行。APP 应用的个数直观地反映了公众参与数据治理的积极程度。

##### 4.4.2 评分量

用户的评分是直接体现了用户对数据提供者以及

数据本身的一种使用体验，是用户对数据质量等内容是否满意最直接的体现。评分量体现了数据应用的质量，从某种程度上反映了数据开放是否真的提高了公民参与社会治理的热情以及社会治理的效率。报告中的评分量主要是以平均评分来进行衡量。

##### 4.4.3 浏览量与下载量

浏览量和下载量从某种程度上体现了公众对于政府数据的关注程度和用户在政府开放数据平台上的参与程度，主要是用户对数据开放后的行为表现。浏览量是指用户对平台数据集的访问情况，下载量是指用户对平台数据集的下载情况。数据的浏览量和下载量的多少会影响用户对数据的关注和利用效果，报告中主要以浏览量、下载量这两个指标来评估当前已经开放的数据集的利用情况。为衡量各地政府开放数据平台数据集的整体关注程度，报告中引入平均浏览量和平均下载量，用平均浏览量和平均下载量来衡量数据集的浏览、下载情况，即将各平台数据集的总浏览量、总下载量除以该平台的数据集个数。通过抓取上海市、广东省、贵阳市、广州市和哈尔滨市 5 地政府开放数据平台数据，统计开放数据集的浏览量、下载量，并计算其平均数据集浏览量和平均数据集下载量，结果如表 4-5 所示。

## 利用效益分析

表 4-5 各地政府开放数据平台数据集的浏览量与下载量情况

城市	数据集个数 (个)	浏览量 (次)	下载量 (次)	平均浏览量 (次)	平均下载量 (次)	APP 应用数 (个)	评分量 (分)
上海市	1364	5077130	1415883	3722	1038	89	1660
广东省	2731	6345786	744248	2324	273	60	24.17
贵阳市	2325	1420282	600996	611	258	15	121
广州市	1237	656903	80151	531	65	8	2552.5
哈尔滨市	1062	347134	109695	327	103	4	89

上海市政府开放数据平台所开放数据集的个数虽然较少，但其在浏览和下载方面均高于其余四个平台，表明其开放的数据集满足用户的需求，从而激励民众进行获取和利用；而哈尔滨市政府开放数据平台无论

是开放的数据集个数，还是在浏览和下载方面都比较少，说明其开放的数据集质量不高，主题内容未能满足用户的需求，同时对开放数据的宣传推广等方面也做的不到位，其数据利用行为还需进一步激发。

# 利用案例

## 5 利用案例

目前针对政府开放数据的利用案例还不够深入，仍需要进一步思考、探索，以促进政府开放数据更好地投入实际应用。本节将通过政府、公众、企业的利用案例，介绍 2018 年政府开放数据的利用现状，以期为促进政府开放数据的开放利用提供借鉴与参考。政府开放数据使用对象主要有政府、公众、企业等类型。目前政府开放数据的利用仍以发布开放数据资源为主，鼓励政府、公众、企业等进行创新，从开放数据中挖掘其价值，指导生产生活实践，同时保障政府开放数据使用对象自身的知情权、参与权、监督权和表达权。政府开放数据方兴未艾，仍有巨大的发展空间，其蕴含的价值也有待开发，各方应携手合作、共同挖掘、互惠互利，为政府开放数据的蓬勃发展贡献力量。

### 5.1 政府层面

相较于公众和企业，政府在政府开放数据中的角色既是数据开放者，也是数据利用者。政府对于政府开放数据的重视与利用，能有效促进政府信息公开工作服务水平，不断拓展政府开放数据的开放范围、增强公开实效。政府在政府开放数据中是强有力的领导者、全方位的政策制定者、决定性的数据供应者、游刃有余的协调者、不可或缺的推进者，也是有重要影响力的关键用户。<sup>1</sup>

### 案例 1：雄安数字孪生城市建设

#### (1) 案例背景

政府开放数据是雄安数字孪生城市建设中必不可少的一环，政府各部门需开放大量数据促进城市治理，优化建设方案，实现科学决策。2018 年 4 月 20 日，《河北雄安新区规划纲要》明确提出建设数字孪生城市，报告指出要坚持数字城市与现实城市同步规划、同步建设，适度超前布局智能基础设施，推动全域智能化应用服务实时可控，建立健全大数据资产管理体系，打造具有深度学习能力、全球领先的数字城市。

#### (2) 利用实例

2018 年底，雄安新区已完成全域 5G 网络覆盖；完成雄安国家级互联网骨干节点建设；建成“双千兆”网络和覆盖全区的窄带物联网，落地示范 5G+VR 360° 景区旅游、无人机“天地一体化”生态监测、智慧护林等，并开展了智慧停车、智慧灯杆等物联网应用示范。根据国家战略部署要求，积极开展 5G 网络规划建设，并将其与现代农业、工业制造业、现代服务业相融合，全面融合政府开放数据，致力于将雄安新区打造为具有高精度城市信息、全域智能设施布局、安全高效智能专网以及智能城市大脑操控的数字孪生城市。将来预计将以未来城市、5G 应用、量子、北斗、自主可控数字城市等联合实验室为依托，以联通 5G 网络为支撑，聚合产业链上下游合作伙伴，持续开展新技术、新业务的创新应用实践，联合开发利用政府开放数据与企业数据，助力雄安新区数字孪生城市建设。

<sup>1</sup> 夏义堃,丁念.开放政府数据的发展及其对政府信息公开活动的影响[J].情报理论与实践,2015,38(12):1-6+19.

# 利用案例

## (3) 结论与建议

雄安新区数字孪生城市建设无法脱离政府开放数据的支撑，政府需制定更为详细的政府开放数据政策，进一步扩大政府开放数据的范围、提高政府开放数据的质量。政府开放数据对于雄安新区提高城市管理科学化、精细化水平，建设高质量、高水平的社会主义现代化城市，具有不可估量的意义与价值。

## 5.2 公众层面

公众作为政府开放数据最广泛的利用群体之一，可通过利用政府开放数据行使监督权与进行科学决策。由于政府开放数据的利用门槛较高、要求较高的数据抽象能力，公众难以进行复杂的数据分析，故通常会选择处理较为简单的数据信息或使用经过加工的二次信息。公众中存在非营利组织和科研群体等较为特殊的用户，此类用户规模较小，但也致力于开放政府数据的生态建设。而从事政府开放数据相关研究的科研人员，因为自身科研工作的需要，往往会较多使用到政府开放数据，对其进行深入探究与思考，挖掘隐藏在数据背后的巨大价值，为保障政府开放数据持续、有效利用提供参考与建议。

### 案例 2：“数游长三角”全域旅游大数据平台

#### (1) 案例背景

从 2015 年至今，上海连续四年举办 SODA 开放数据创新应用大赛，凝聚了众多大数据人才，推动了大数据产业的发展。2018 年，江浙皖三省加入大赛并首次开放长三角数据，以“数联长三角众创新生活”为

主题，充分利用长三角丰富的开放数据资源、雄厚的大数据产业基础，以及扎实的产学研基础及人才，服务长三角更高质量一体化发展。主题为“数联长三角众创新生活”，旨在凝聚社会各界大数据创新力量，搭建开放数据创新应用、跨境数据发展的交流合作平台，聚焦推进长三角大数据技术发展，共话大数据产业前沿。

#### (2) 利用实例

由上海对外经贸大学工商管理学院师生参与的作品《“数游长三角”全域旅游大数据平台》，集合了技术公司、高校科研和政府应用三方力量，综合运用十多种大赛数据及其他公开数据，进行产品设计、算法研发和数据挖掘，经过激烈角逐，最终获得此次大赛“优秀奖”。工商管理学院师生研发设计的“数游长三角”全域旅游大数据平台融合 10 种不同的数据资源，进行政府大数据与其他社会大数据整合处理分析从而实现政府开放数据开发利用，最终提供 3 款产品和服务，包括为政府提供“长三角旅游全景监控中心”；为景区提供“景区实时检测服务”；为公众提供“长三角旅游推荐”。此产品有助于实现数据共享的社会价值，打造长三角全域旅游大数据产品，为公众假期出行旅游提供便利。

#### (3) 结论与建议

SODA 开放数据创新应用大赛是政府聚集公众之力，以“数据众筹、应用众包、问题众治”为理念，他山之石用以攻玉，面向全社会征集大数据解决方案，是较为成熟的政府开放数据利用大赛模式。此类比赛联

动政府、企业与社会公众，促进政府开放数据的价值挖掘与开发利用、提高政府开放数据利用效率，值得其他政府部门对于政府开放数据的利用进行研究、思考与借鉴。

### 5.3 企业层面

企业作为具有较大政府开放数据需求与较强数据分析的利用者之一，对于促进政府开放数据具有极其重要的意义。企业可将开放政府数据应用于经营管理的各个环节，有助于企业进行产品预测、科学决策、制定发展方向，提高企业的社会与经济效益。通过对政府开放数据有效利用，企业可以在获得经济利益的同时间接服务公众，从而有效地促进政府开放数据运动的发展。然而，目前的研究更局限于政府与企业之间的合作与共享，更多样化的协同利用模式仍然很少<sup>1</sup>。企业所拥有的较为强大的技术能力，能充分发挥政府开放数据的经济和社会效益。

#### 案例 3：阿里 ET 城市大脑

##### （1）案例背景

2018 年 11 月 15 日，央视以《坚持发展“枫桥经验”——中国基层社会治理现代化之路》为题的专题片《智惠于民》，报道了大数据、人工智能等新技术在提升社会治理现代化方面的进展，其中城市大脑协助衢州、余杭等地推进科学决策、高效服务得到了高

度评价，成为新时代“枫桥经验”的延伸<sup>2</sup>。城市大脑将用于解决杭州交通拥堵等问题，目标是依据城市交通体检、城市警情监控、城市交通微控、城市特种车辆、城市战略规划等场景需求，让数据帮助城市来做思考和决策，将杭州打造成一座能够自我调节、与人类良性互动的城市。

##### （2）利用实例

余杭城市大脑作为智能中枢，汇聚大量物联网设备、视频设备数据，可指挥全区 6 个消防队、10 个专职消防队、60 个微型消防站，以及 120、110、水务、电力等多个联动部门。2018 年 9 月 17 日上午，杭州余杭消防大队进行了一次特别的演练：上午 10 点 57 分，阿里云 ET 城市大脑报警，汤加锦绣社区星韵北路有火情，需派出 1 辆水罐消防车救火，同时向 120、110、水务、电力等联动部门推送信息，消防车、救护车等立刻出发进行救援。此次行动联合各政府部门，通过对政府数据的综合分析处理实现警情预判与特种车辆调配，体现了阿里 ET 城市大脑服务的智能性与高效性，表明了政府开放数据中蕴含的巨大价值。

##### （3）结论与建议

阿里 ET 城市大脑整合政府在交通运输过程中产生的大数据，将其应用于城市交通体检、城市警情监控、城市交通微控、城市特种车辆、城市战略规划等实际场景中，结合企业强大的交通大数据分析能力，为城

1 汤志伟,郭雨晖.我国开放政府数据的利用:基于 CNKI 的系统性文献综述[J].情报杂志,2018,37(07):176-181+65.

2 阿里云 ET 城市大脑的央视 7 分钟:人工智能可以深度参与城市治理 [EB/OL].[2018-11-16].<https://baijiahao.baidu.com/s?id=1617289840040583116&wfr=spider&for=pc>

# 利用案例

市交通治理提供新的解决方案，提高城市交通运行效率。

大数据时代，政府开放数据的数据量激增，海量数据价值密度相对较低，需要通过商业运作方式开发

政府开放数据价值，提升政府开放数据利用效率，从而倒逼政府开放数据的发展，以实际利益促进政府开放数据运动。

# 结论与对策

## 6 结论与对策

### 6.1 总结

#### (1) 数据获取平台状况

2018 年，政府开放数据平台建设数量呈现不均衡特征，平台类型多为集中专有式。华东地区和华南地区的政府开放数据平台建设情况较好，并呈现出将建设重点转移到市级平台的趋势，这一趋势在华东地区表现最为明显。华中、东北、西南、西北地区平台建设情况较为缓慢。从平台上线数量来看，2018 年我国新增政府开放数据平台 36 个，平台上线数量同比增长 138.46%，当前政府开放数据平台建设呈现出高速发展态势。从平台类型来看，2018 年我国政府开放数据平台大多为专有式类型，即数据统一汇聚在一个专门的平台上进行开放，仅浙江、梅州、阳江、肇庆地区的政府开放数据平台类型为嵌入式。从平台所属地区来看，2018 年我国政府开放数据平台主要集中分布于华东、华南地区。华东地区在全国范围内处于领先地位，市级平台建设速度较快，副省级、省级平台建设速度较缓。华南地区市级平台建设较省级、副省级平台具备连贯性，但建设速度较为平缓，没有大幅增长。

#### (2) 可供分析数据状况

开放数据是用户进行分析处理的原材料，其质量直接影响开放数据的利用效果。目前可供分析数据的基本状况如下。

各地平台开放的数据集数、接口数与应用数呈现不均衡分布，数据应用数量较少。目前贵阳、广东、

武汉所开放的数据集数量排在前三位，开放数据集基数较大，但所开放的接口数占比极小；相比之下，济南和哈尔滨开放的数据集数量虽少，但所开放的接口数较多。此外，上述平台所开放的的应用数量远远低于数据集数量和接口数量。

数据集数量规模化已有显现，但数据主题覆盖情况不一，数据类型分布呈不均衡状态且无五星级数据。现有数据集在法律服务、宗教信仰等数据主题上存在明显的缺失。在数据资源格式上，各地区政府开放数据类型的分布呈现不均衡状态，广东、上海、武汉、贵州的开放数据格式较为丰富，能更好地满足不同需求的数据使用者更方便快捷地对数据进行增值利用，但根据 Berners — Lee 提出的开放数据五星标准，样本区域内均无五星级数据。

#### (3) 利用行为综合分析

综合来看，用户对不同主题的数据集利用存在差异。对经济工商、教育科技、社保就业和交通出行类数据等与民生相关的数据的利用程度较深，利用效率较高。经济工商类数据是企业用户关注和需求的重点，而医疗健康类数据则是全社会关注的热点。相比之下，用户对能源环境、公共安全等社会治理领域的问题目前关注不多。

#### (4) 利用综合效益

政府开放数据利用效益的评价研究是一个动态变化的过程。政府开放数据利用效益与数据可用性的关联程度最高，是政府开放数据利用效益评价中最重要

# 结论与对策

的一级指标。数据集的浏览量和下载量具有较强的相关性，在对政府开放数据利用效益的评估中也起到了很重要的作用。不同地方政府的利用效益具有明显的层次性，上海的开放数据利用效益最好，接下来依次是广东省政府开放数据、贵阳市政府开放数据、广州政府开放数据。

## 6.2 存在问题

### (1) 数据获取平台

① 违背平台集约化建设理念。目前，政府开放数据平台呈现不均衡的发展趋势，市级数据开放平台开始被重点建设，这与平台集约化建设理念有所冲突，且不利于数据的纵向流通和共享。以省级或副省级平台作为基础性平台在一定程度上即可满足政府开放数据的需求。

② 平台区域分布不均衡。数据开放平台区域分布不均衡，集中分布于华东、华南地区，华中、东北、西南、西北地区开放平台数量较少。华东地区的平台建设在全国范围内处于领先地位，华南地区市级平台建设较省级、副省级平台具备连贯性，但建设速度较为平缓。

③ 平台功能有待提升。用户进行数据资源获取时，存在数据开放平台浏览界面不够友好、检索方式单一且搜索功能不够精确、排序方法简单以及难以批量下载的问题。数据开放平台在检索结果页面无法按照用户的个性化需求进行排序，大多数的平台都尚未提供批量下载的功能，降低了用户获取数据的效率。

### (2) 可供分析数据

开放数据的质量会对用户数据获取和利用产生影响。当前，政府开放数据存在开放主题及开放机构不够充分，数据可持续性不高、开放度相对较低、主题分类差异较大、筛选过程质量有待提高等问题。

① 开放主题及开放机构不够充分。政府开放数据资源整体存在主题及开放机构不够充分的问题。政府开放数据在科技、交通、经济、教育、医疗等方面的主体涉及较多，而在法律服务、宗教信仰等数据主题上存在明显的缺失；在开放机构上，主要涉及各级政府机构和公共企事业单位，但尚未涵盖政府的所有职能如国防部门等。开放的数据主题领域较少，数据类型分布尚不均衡，更多对用户有价值的数据亟需开放。

② 数据可持续性不高。在开放数据的可持续性方面，目前存在数据更新频率慢、更新的规范化程度不够高的问题。较多数据平台的数据更新频率较低或者从未更新，数据老化情况严重。数据更新的规范化程度不够高主要体现为部分数据集更新的元数据标准不够清晰，存在更新后数据字段重复的问题。

③ 开放度相对较低。在数据的开放度方面，存在开放格式不充分、用户尚不能完全免授权获取与自由使用数据等问题。我国政府开放数据平台提供的数据格式较为单一，且每一种数据集所选择的供应格式也比较单一。大多数的数据在获取时需网站注册并进行实名认证，用户不能完全自由使用数据。

④ 主题分类差异较大。各地的政府开放数据平台使用的主题分类存在差异，缺乏统一的主题分类标准。各个平台都有一套自身的数据主题分类标准，从而使

得各个平台相互孤立，一定程度上阻碍了地区之间政府开放数据的共享与联通。

⑤ 数据筛选过程质量有待提高。各主题开放数据整体与部分的特征并不具有一致性，在整体浏览率较高的数据集中，并非任意一条数据的浏览率都高。因此，在开放数据前期的筛选过程中，针对某一主题的数据，并不能只追求数据的全面性，也需要有一定的针对性，从而在利用过程中减轻用户的数据筛选负担。

#### （3）数据利用频度低，公众利用意识不强

综合数据集的浏览量和下载量分析可以发现，排名靠前的主题数据的浏览量和下载量也未超过百万，在我国的人口基数条件下，开放数据利用的用户数据所占比例较低，数据利用频度较低，这在一定程度上反应了全民的开放数据利用意识不强，有待进一步提高。

#### （4）数据利用广度窄且深度浅

用户对不同主题领域数据的利用情况存在差异，某些主题领域的的数据价值有待挖掘。用户利用的数据主要集中在经济工商、教育科技、社保就业和交通出行领域，而对能源环境、公共安全等社会治理领域的问题目前关注不多，总体的数据利用广度较窄。此外，数据浏览情况因地而异，各主题数据下载率分布不均衡，整体的数据利用深度较浅，用户对经济工商、教育科技、社保就业和交通出行类数据的利用程度相对深入，其他主题领域的的数据价值有待进一步挖掘。

#### （5）数据利用效益不高

我国还没有形成政府开发数据的完整产业链，产业链的上下游尚未形成协作发展模式。企业参与政府开放数据的开发的盈利模式与服务方式还不明确，缺乏大规模的能够带动政府数据开发产业发展的龙头企业，由公众需求拉动和企业主导开发的政府开放数据共享、开发利用的生态还没有形成，政府开放数据的整体利用效益不高。

### 6.3 对策建议

#### （1）提升政府开放数据利用意识

积极提升社会的开放数据利用意识。一方面，政府开放数据利用的意识有待加强，政府作为开放数据的主体，数据开放在一定程度上打破了原有政府信息资源开发利用框架，数据作为原材料的免费提供，使得政府部门难以垄断政府数据资产价值的创造，以往成本回收模式和边际成本模式难以持续运行，以及因担心开放后的安全问题，部分政府机关从思想上对数据开放存在抵制情绪。可以通过上级政府的强制要求和建立激励机制进行调节，同时也要加强数据开放的教育工作，提升政府自身数据开放的意识。

另一方面，用户利用政府开放数据的习惯还未养成。用户产生信息需求时，总是倾向于寻找经过加工的资料，而忽略了具有高价值的政府开放数据。政府可在政策以及各项活动中积极倡导对开放数据的应用，并融合在公共生活的方方面面，从而使公众在社会活动中能意识到开放数据的价值与必要性。

#### （2）优化政府开放数据质量

# 结论与对策

政府机构在发布数据前应建立和完善质量保障机制，通过对数据源的管理、设立数据的标准规范，开放前的数据清洗等方式提升政府开放数据的质量。另一方面，用户在使用数据前应进行数据的质量检查工作，发现问题可及时向数据平台进行反馈。在管理方面，应结合政府开放数据的现状，建立数据质量管理流程，提升质量管理的科学性。同时，政府自身或借助第三方机构进行政府开放数据质量的评价与考核，为优化政府开放数据质量提供支撑。

### (3) 促进政府开放数据价值释放

政府开放数据利用的目的在于其价值的释放，而

这与用户的数据利用能力息息相关。政府应加强对数据用户的培育，提升其数据利用技能。相关部门可以组织专家学者组成数据开放利用培育团队，通过公开课、新媒体等渠道推进数据利用技能培训和教育，为开放数据用户提供各类指南、工具使用手册等。对于企业用户，培育带动政府数据开发产业发展的龙头企业，明确其开发利用政府数据的服务方式和盈利模式，充分发挥政府开放数据的价值。

### (4) 提高数据利用率，做到“三好三可”

政府开放数据重在对数据的利用，提高数据利用率，做到“三好三可”。

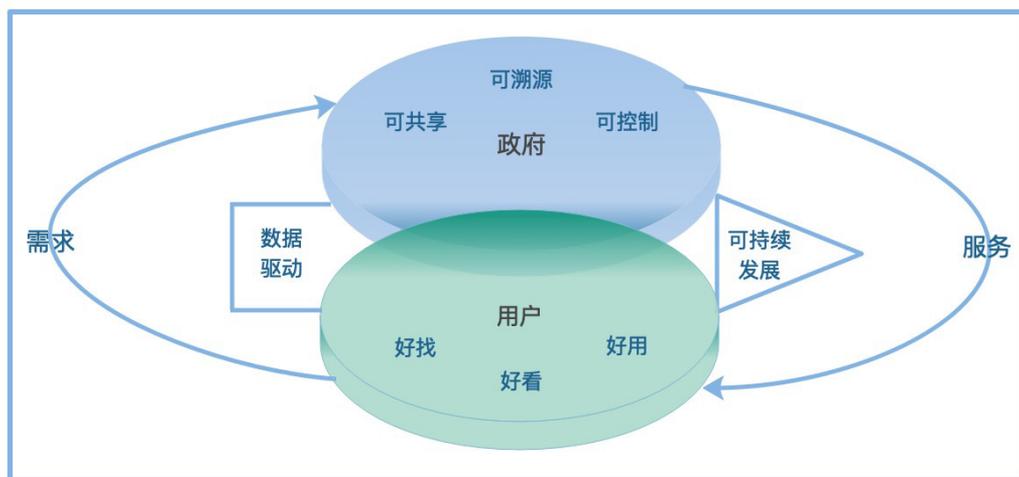


图 6-1 政府开放数据利用“三好三可”标准

#### ① 好找

好找体现在用户不仅能够找得到政府开放数据，而且可以找得快、找得准。政府开放数据平台是用户获取开放数据的主要途径，其设计应做到界面简洁明了，功能齐全、信息分类合理，要准确地传递数据信息，

使用户可以精准获取信息内容，避免出现信息误解。尽量使用户在不需要学习的情况下就可以便捷地开展平台操作；检索方面，应丰富平台的检索方式和排序方式，满足用户在特定情境下的个性化检索需求。同时，政府部门应逐步加强数据开放平台的人性化设计，

减少用户注册和下载数据过程中的繁琐操作。相关部门应加大技术方面的投入，实现网站对于用户的快速响应，确保用户在获取开放数据时找得到、找得快且找得准。

## ② 好看

好看是使政府开放数据具备较好的视觉体验，既能看得见，又能看得懂、看得舒适。视觉信息比文本信息更易于接收，数据可视化可以促进用户快速接收信息，在一定程度上增强了政府开放数据对用户的吸引力。首先，要确保为用户提供看得见的政府开放数据，逐步加强政府开放数据的可视化建设。以图、表的形式展示数据动态变化信息，同时可考虑创造性地利用现代通信工具、多媒体技术或 3D 虚拟现实信息搜索环境等先进的技术来加强计算机生成的虚拟环境的浸入感，使用户获得更好的体验。其次，政府开放数据可视化的界面要简洁清晰，确保用户看得懂展现的数据内容。同时，数据可视化的设计要尽量符合美学的设计原则，为数据用户带来舒适的视觉体验。

## ③ 好用

好用主要体现在用得着、用得顺和用得有效三个方面。首先，政府开放的数据应是用户关切或能满足用户信息需求的数据。政府部门应在确保信息安全的情况下开放多领域多主题的数据，并及时更新，保证数据的完整性和全面性。同时，要加强政府各部门以及不同行政层级政府间的统筹协作，制定各类数据发布、更新标准，公布数据发布日程以及修改说明等。其次，用得顺体现在政府开放数据提供的数据格式可

以方面用户进行后期的数据处理，政府在开放数据时应注重提供丰富的数据开放格式。利用政府开放数据的最终目标是能够满足用户需求并产生价值，因此政府可以通过提供有用且高质量的数据来确保用户利用数据后产生好的效果。

## ④ 可共享

可共享的重点在于数据开放共享的标准和规范设立。目前我国对数据没有统一的标准，数据口径不同和格式多样等都阻碍了数据共享，造成了一个个独立的信息孤岛仅实现内部共享，无法发挥数据的整体优势。政府应加快建立一套统一的、法定的数据开放共享标准规范，明确开放数据的技术标准和管理标准，同时建立健全政府开放数据共享机制，强化部门协同，以目标为导向，按需共享，规范建设过程，保证共享数据的质量，确保数据共享的安全。

## ⑤ 可溯源

数据可溯源是评估开放数据质量和可信度的基础。要加强政府开放数据溯源元数据建设，确保政府开放数据有明确的数据来源、路径和时间，在元数据中设立明确的相关字段进行标注。目前的政府开放数据元数据中含有相应的溯源信息，但总体质量还不高，溯源能力较弱。因此，需要进一步加强基于全生命周期的政府开放数据利用溯源元数据体系建设。

## ⑥ 可控制

建立对政府开放数据利用全过程的管理控制机制，从制度、技术和舆论三个维度对数据开放利用过程提

## 结论与对策

供全方位的保障。制度维度，一方面政府应不断出台政府开放数据利用相关政策，并加强政策的推行力度；另一方面，要建立科学合理的开放数据利用管理制度，明确各过程的管理方式和各部门的管理职责。技术维度，理清数据开放利用的业务流程，在数据筛选、开放、

共享、下载等各个环节加强控制，明确各环节的技术规范。此外，政府应加强开放数据利用的舆论引导，鼓励公众参与到开放数据利用的管控过程中，实现开放数据利用的共治共建共享。

# 附录

## 附录

### 1.1 样本选取与数据采集

### 1.2 工具选取与研究方法

## 联合发布方

---



华中师范大学信息管理学院



湖北省数据治理与智能决策研究中心

## 报告出品团队

---

段尧清 夏立新 李玉海 娄策群 易 明  
陈 玲 姚 兰 刘宇明 何思奇 王冰清  
林 平 周 密 尚 婷