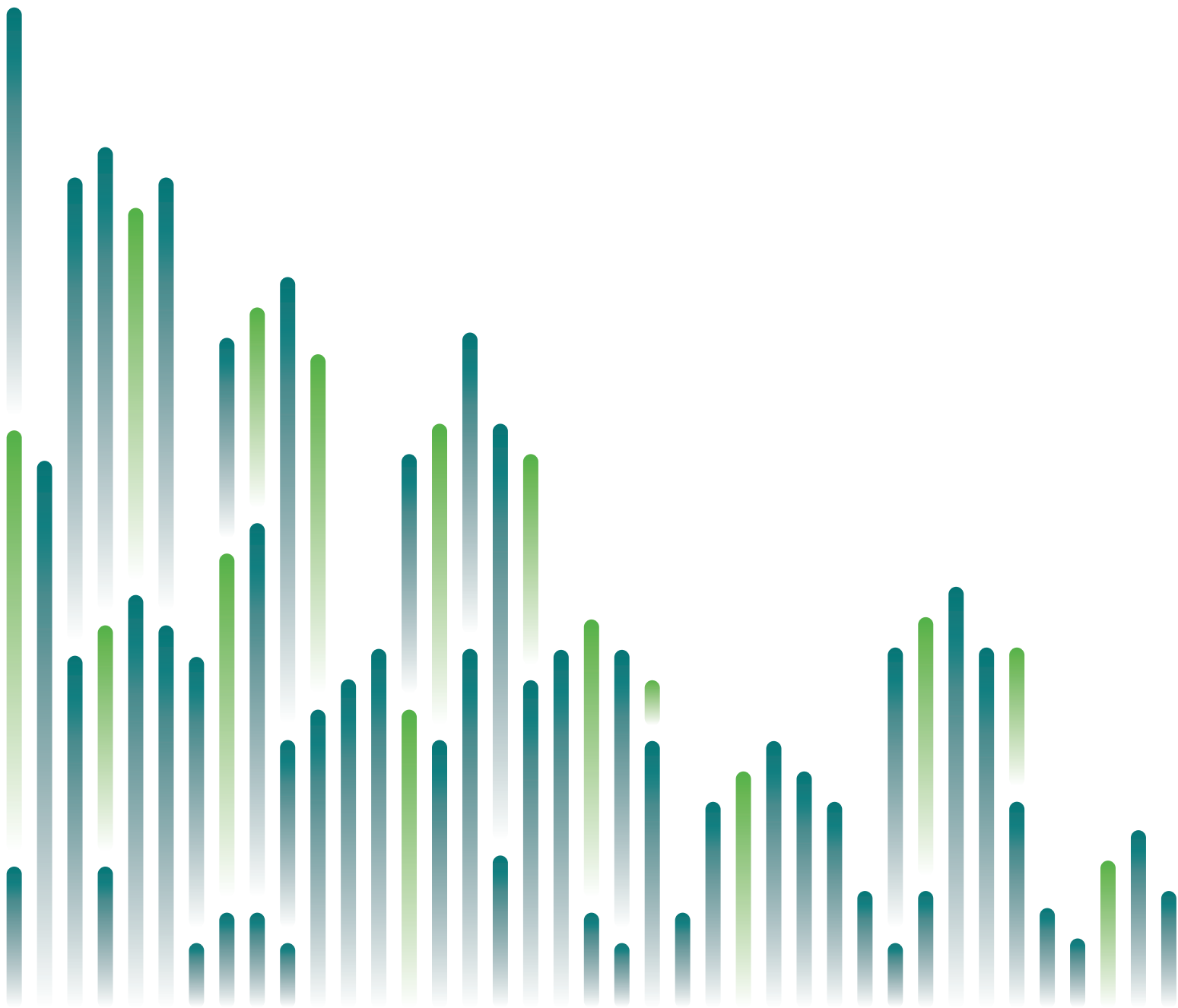


中国开放数据指数
CHINA OPEN DATA INDEX

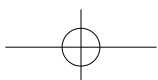
中国地方政府数据开放平台报告 | 数据获取

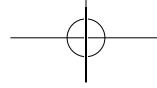
开放数据 蔚然成林

1
2017.5



联合发布方 复旦大学 & 提升政府治理能力大数据应用技术国家工程实验室 指数出品方 复旦大学数字与移动治理实验室





CHINA OPEN DATA INDEX

中国地方政府数据开放平台报告

数据获取

开放数据 蔚然成林

1
2017.5



目录 INDEX

- 03 一 背景与趋势
- 04 二 评估方法
 - 04 2.1 定义、原则与标准
 - 04 2.2 评估范围
 - 06 2.3 指标体系
 - 07 2.4 数据采集及分析
- 08 三 主要发现
 - 08 3.1 总体分布
 - 10 3.2 开放数据集总量
 - 11 3.3 数据开放度
 - 16 3.4 数据覆盖面
 - 24 3.5 数据持续性
- 27 四 “开放数林” 指数分值与排名
 - 27 4.1 各项指数均值
 - 27 4.2 指数分值与排序
- 30 五 结论与建议



1. 背景与趋势 Background

一、背景与趋势

数据是国家的战略性资源，一个国家有大量基础性、关键性的数据掌握在各级各地政府部门手中。大数据建立在开放数据的基础上，政府部门在履行行政职责过程中制作、获取和保存了海量的数据资源，这些数据是整个社会的公共资源，应取之于民，还之于民，释放数据能量，创造社会价值。在保障国家秘密、商业秘密和个人隐私的前提下，将政府数据最大限度地开放给社会进行开发利用，将有利于加强政府透明度，激发创新活力，提高服务水平，转变经济发展方式，提升政府治理能力。

近年来，开放政府数据已在全球范围内迅速推进。根据《2016 联合国电子政务调查报告》的数据，截止 2016 年，联合国 193 个成员国中已有 106 个提供了开放数据目录。我国政府也高度重视开放政府数据。2015 年 8 月，国务院印发的《促进大数据发展行动纲要》要求稳步推动公共数据资源开放，加快建设国家政府数据统一开放平台。2017 年 2 月，中央全面深化改革领导小组第三十二次会议审议通过了《关于推进公共信息资源开放的若干意见》，要求着力推进重点领域公共信息资源开放，释放经济价值和社会效应。2017 年 5 月，国务院办公厅又印发《政务信息系统整合共享实施方案》，明确要求“推动开放，加快公共数据开放网站建设”。

2012 年以来，我国已有近二十个地方政府陆续推出数据开放平台，先行先试，积极探索，积累了许多经验，也仍

然面临着许多挑战。我国各级城市是公众集中生活的重要空间，城市数据具有高密度高价值的特点，开放城市政府数据有利于吸引社会各方基于城市的实际需求和应用场景对开放数据进行融合利用，创造社会经济价值，并形成正向反馈，进一步推动政府数据开放，构建起一个动态循环的开放数据生态系统。在此背景下，复旦大学与“提升政府治理能力大数据应用技术国家工程实验室”联合发布国内首个地方政府数据开放报告暨中国开放数林指数。

报告由复旦大学数字与移动治理实验室基于多年研究积累和数据监测分析制作完成。基于开放数据的原则与标准，借鉴多家国际性评估报告实践，立足我国政府数据开放的实际情况和发展阶段，报告构建起一个系统科学、多维度、可操作的评估框架，并基于该框架对我国现有的地方政府开放数据平台进行综合评价，制作指数报告，提出优化建议。

开放数据，蔚然成林。一棵棵开放“数木”的丛然共生、成荫如盖、直至叶茂花开，终将成长为一片繁盛多样、平衡稳定、循环持续的“开放数林”。

《中国地方政府数据开放平台报告》将定期发布“开放数林指数”，持续追踪我国地方政府数据开放工作的进展与问题，精心测量“数木”们的形态、密度、材质与价值，助力中国开放数据生态系统的形成与发展。

2. 评估方法

Methodology

二、评估方法

2.1 定义、原则与标准

2007年12月，30位开放数据倡导者聚集在美国加利福尼亚州，首次提出了开放政府数据的8大基本原则：完整的、原始的、及时的、可公开获取的、可机器读取的、非歧视获取的、非专属的、免于授权许可的。会议还将“数据”定义为“一切以电子形式存储的记录”，并指出公共数据是指不涉及隐私、安全和其他特定限制的数据¹。

《2016联合国电子政务调查报告》将开放政府数据定义为“主动在网上公开政府信息，使任何人都能不受限制地获取、再利用和再分发”²。根据世界银行的定义，开放数据是“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，并最大程度保持原始出处和开放性的数据”。所谓“开放性”，应具备两个维度的特性：一为法律性开放，即这些数据必须被置于公共领域，或处于自由利用条款下，受到最低程度的限制；二为技术性开放，即数据必须以机器可读的、最好是非专属的、电子格式发布，从而使任何人可使用通用的、免费的软件获取和利用这些数据，数据还应被放置于公共服务器上，可被公开获取，不受密码或防火墙的限制³。

2015年，《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、在任何时间和地点进行自由利用、再利用和分发的电子数据。该宪章还提出六项原则：以开放为默认原则、及时和全面、可公开获取和可利用、可比较和互操作、致力于改善治理与公众参与、

1 www.opengovdata.org

2 <https://publicadministration.un.org/egovkb/en-us/reports/un-e-government-survey-2016>

3 <http://opendatatoolkit.worldbank.org/en/essentials.html>

致力于包容性发展和创新⁴。

纽约大学政府实验室（Govlab）对国际上具有代表性的十一个研究机构、评估指标、政府部门和咨询公司界定的“开放数据”定义进行梳理后发现，最多被提及的开放数据标准包括免费、公开提供、非排他性、可利用结构、开放授权和可再利用等⁵。

2017年，国务院办公厅印发的《政务信息系统整合共享实施方案》指出要向社会开放“政府部门和公共企事业单位的原始性、可机器读取、可供社会化再利用的数据集”。国家公共数据开放的有关要求也指出要保证开放数据的完整性、准确性、原始性、机器可读性、非歧视性、及时性，方便公众在线检索、获取和利用。

本次评估根据以上定义、原则与标准，基于目前我国政府数据开放的实际情况和发展阶段，确定了报告的评估对象、指标体系、数据采集分析方法、以及指数计算方法。

2.2 评估范围

报告根据公开报道，以及使用“数据+开放”、“数据+公开”、“政务+数据资源”、“政府+数据服务”、“地名+政府数据”、“地名+数据”等关键词进行搜索，发现了截至2017年4月我国已上线的政府数据开放平台，并将符合以下条件的开放数据平台纳入评估范围：

(1) 平台域名中出现 gov.cn，作为确定其为政府官方认可的数据开放平台的依据。

(2) 平台形式为“统一专有式”或“统一嵌入式”。“统

4 <http://opendatacharter.net/>

5 <http://odim pact.org/resources.html>

一专有式”是指开放数据统一汇聚在一个专门的平台上进行开放，“统一嵌入式”是指开放数据统一汇聚为一个栏目板块，嵌入在政府门户网站或政务服务网站上。各个条线部门建设的非集中式开放数据的平台不在本报告的评估范围内。

(3) 平台所代表的地方政府的行政级别一般为地级以上。需要特别说明的是，作为国内最早开放数据的区级政府，佛山市南海区在国内政府数据开放领域进行了先行探索，为持续跟踪比较，“数说南海网”作为唯一的区级平台被纳入此次评估的范围。

(4) 平台上确实开放了电子格式的、可通过下载或接口形式获取的、结构化的数据集。即使名为“数据开放”平台，但实质上只提供了非结构化的文本内容，不存在可通过下载或接口形式获取的、结构化的数据集，更多属于传统“信息公开”性质的平台未纳入本次评估范围，如江西省人民政府门户网站“用数据”模块、四川省人民政府网站上的“开放数据”模块和广东清远市人民政府网“数据开放”频道等。

基于以上选择标准，被纳入本次报告评估的地方政府数据开放平台共19个，这些地方的平台符合政府数据开放的

表 1 评估范围

	平台名称	地点	层级	平台域名
1	北京市政务数据资源网	北京市	省级	www.bjdata.gov.cn
2	开放广东数据服务网	广东省	省级	http://www.gddata.gov.cn
3	贵州省政府数据开放平台	贵州省	省级	http://www.gzdata.gov.cn
4	上海政府数据服务网	上海市	省级	http://www.datashanghai.gov.cn
5	浙江政务服务网	浙江省	省级	http://data.zjzfw.gov.cn
6	广州市政府数据统一开放平台	广东省广州市	副省级	http://www.datagz.gov.cn
7	哈尔滨市政府数据开放平台	黑龙江省哈尔滨市	副省级	http://data.harbin.gov.cn
8	青岛市政府数据开放网	山东省青岛市	副省级	http://data.qingdao.gov.cn
9	深圳市政府数据开放平台	广东省深圳市	副省级	http://opendata.sz.gov.cn
10	武汉市政府公开数据服务网	湖北省武汉市	副省级	http://www.wuhandata.gov.cn
11	长沙数据开放	湖南省长沙市	地级	http://data.changsha.gov.cn
12	数据东莞网	广东省东莞市	地级	http://dataopen.dg.gov.cn
13	佛山市数据开放平台	广东省佛山市	地级	http://www.fsdta.gov.cn
14	贵阳市政府数据开放平台	贵州省贵阳市	地级	http://www.gyopendata.gov.cn
15	梅州市人民政府数据开放平台	广东省梅州市	地级	https://www.meizhou.gov.cn/opendata
16	无锡市政府数据服务网	江苏省无锡市	地级	http://etc.wuxi.gov.cn/opendata
17	肇庆市人民政府“用数据”模块	广东省肇庆市	地级	http://www.zhaoqing.gov.cn/sj kf
18	湛江市政府数据服务网	广东省湛江市	地级	http://data.zhanjiang.gov.cn
19	数说南海网	广东省佛山市南海区	区级	http://data.nanhai.gov.cn

2. 评估方法

Methodology

基本特征，是我国开放政府数据的先行者。具体平台名称、所属地方政府和平台域名如表 1 所示。

“数据”是开放数据的核心，包括数据数量、种类、标准、质量和价值等方面，因此本次评估主要针对各地平台上开放的数据本身。为聚焦重点，本次评估暂不包括开放政府数据的法制、组织、技术、经济和社会环境；开放数据平台功能和体验；开放数据的使用情况和效果等方面。

报告的主要观察对象为我国地方政府数据开放平台上可通过直接下载或 API 接口两种方式公开获取的、电子形式的原始数据集及其相关信息；不包括平台上的 APP 应用、数据报告、可视化呈现等数据应用产品。基于开放数据的非歧视和公开获取原则，通过内部授权、协议开放等形式开放的数据也不属于本报告的评估对象。

2.3 指标体系

(1) 指标框架

复旦大学数字与移动治理实验室通过对十多个具有代表性的国内外政府数据开放评估项目在“数据层面”的评估指标进行梳理和分析后发现⁶，最受重视的二级指标依次是：数据标准、数据质量、数据管理、数据数量、数据类别、关键数据集等，其中开放授权、机器可读、可获取、非专属、及时、元数据提供等三级指标最受关注。

在充分借鉴国际性开放数据评估报告指标体系的基础上，立足于目前我国各地开放数据实践的实际情况与发展阶段，本着系统科学、多维度、可操作的原则，本报告建构了一套评估指标框架（如表 3 所示）。框架共包括四个一级指标：数据集总量、数据开放度、数据覆盖面与数据持续性。

⁶ 郑磊、关文雯，开放政府数据评估框架、指标和方法研究，《图书情报工作》，第 60 卷第 18 期 <http://www.dmg.fudan.edu.cn/?p=2790>

(2) 指标权重

为确定各个指标的权重，报告出品方邀请 50 多位我国数据科学领域的学者和数据开发利用专业人士组成专家评委会（详细名单见报告附件）对各项评估指标的相对重要性进行排序，以反映第三方学术机构的专业视角和数据利用者们的实际需求。排序结果如表 2 所示，被排在最重要位置的指标是“数据集价值”，这反映了各界对高价值数据集的需求。然而，数据的价值要通过实际应用来展现和评估，由于目前我国政府数据开放工作刚刚起步，大量高价值政府数据仍未开放，数据尚未被充分应用以展现其价值，因此本次报告暂不评估该项指标。被专家评委会排在第二至第五位的指标依次是开放数据格式、元数据提供、开放授权以及数据集更新，而开放重点领域关键数据集、数据集主题覆盖面与数据集总量的重要性相对靠后，这反映了学界和业界更为重视数据标准和数据质量，而反对一味追求数据数量。

表 2 评委专家对评估指标相对重要性排序的结果

指标名称	排序得分	重要性
数据集价值	3.37	1
开放数据格式（技术性开放）	3.66	2
元数据提供	3.70	3
开放授权	3.72	4
数据集持续更新	3.83	5
开放重点领域关键数据集	4.47	6
数据集主题覆盖面	4.51	7
数据集数量	4.51	7

评估指标框架及指标权重的设置意在推进我国各地开放数据实践的持续健康发展，需兼引导性、相关性、明确性、可衡量性、可实现性和动态性，综合考量国际性评估报告的

指标体系、专家评委会的评分结果、以及目前我国各地发展现状、主要问题和亟需引导的方向，本期报告确定了各级指标对应的权重（详见表 3），对数据开放度赋予最高的权重（50%），其次是数据持续性（20%）、数据集覆盖面（15%）与数据集总量（15%）。随着各地政府数据开放实践的深入推进，指标框架及指标权重将进行动态调整。

2.4 数据采集及分析

报告采用自动抓取和人工观察方法采集数据。以 2017 年 4 月 18 日为数据采集始点，在之后四周内的四个时间点持续采集数据，采集时间分别为 2017 年 4 月 18 日、2017 年 4 月 24 日、2017 年 4 月 30 日与 2017 年 5 月 7 日。报

告主体分析部分以 4 月 18 日采集的数据为基准，而在平台数据动态更新部分则使用四个时间点采集的数据进行追踪对比分析。报告对采集到的各项指标的数据进行了描述性统计分析和交叉分析。

2.5 指数计算方法

报告对各地平台在各项指标上按照 0-5 共 6 档分值进行评分，由低到高分别为 1-5 分（“开放授权”指标的分值例外，4 项三级指标共 5 分，每满足一项得 1.25 分），相应数据缺失或完全不符合标准则分值为 0。各地平台的最终指数总分为每个单项指标的分值乘以相应的权重，再换算成百分制。

表 3 指标框架与权重

一级指标 (权重)	数据集总量		数据开放度			数据覆盖面			数据持续性			
	15%		50%			15%			20%			
二级指标 (权重)	20%			20%		10%	5%	5%	5%	8%	6%	6%
	开放授权			技术性开放		元数据提供	主题覆盖率	部门覆盖率	关键词覆盖率	持续增长	动态更新	历史存档
三级指标 (权重)	5%	5%	5%	5%	8%	7%	5%	10%				
	免费获取	非歧视性	自由利用	自由传播与分享	机器可读格式	开放格式	API 接口	基本元数据覆盖率				

3. 主要发现

Key Findings

三、主要发现

3.1 总体分布

(1) 上线时间

自 2012 年上半年上海推出我国大陆地区第一个开放数据平台起，截止到 2017 年上半年，我国陆续上线了十多个符合开放政府数据基本特征的地级以上平台（如图 1）。以 2015 年 8 月底国务院印发《促进大数据发展行动纲要》为分水岭，在此之前，各地开放数据平台增长较为缓慢，上海和北京最早进行了探索，此后无锡、佛山南海、湛江、武汉等地先后上线；《促进大数据发展行动纲要》出台后各地平

台增长速度明显提升，15 年下半年新上线了 4 个平台，16 年下半年又增加了 6 个平台，17 年上半年也陆续有新平台推出。

(2) 地区分布

截止 2017 年 4 月，我国各地上线的符合政府数据开放基本特征的平台地理位置分布如图 2。标注为深色的为省级平台，浅色为地级市。由图可见，我国绝大多数地方政府尚未开放数据。目前上线的平台主要分布在东南部沿海地区，部分分布在西南与华中地区，长江以北分布较少。其中，广东省地级市开放平台分布尤为集中，形成了我国最为密集的省级“开放数林”。

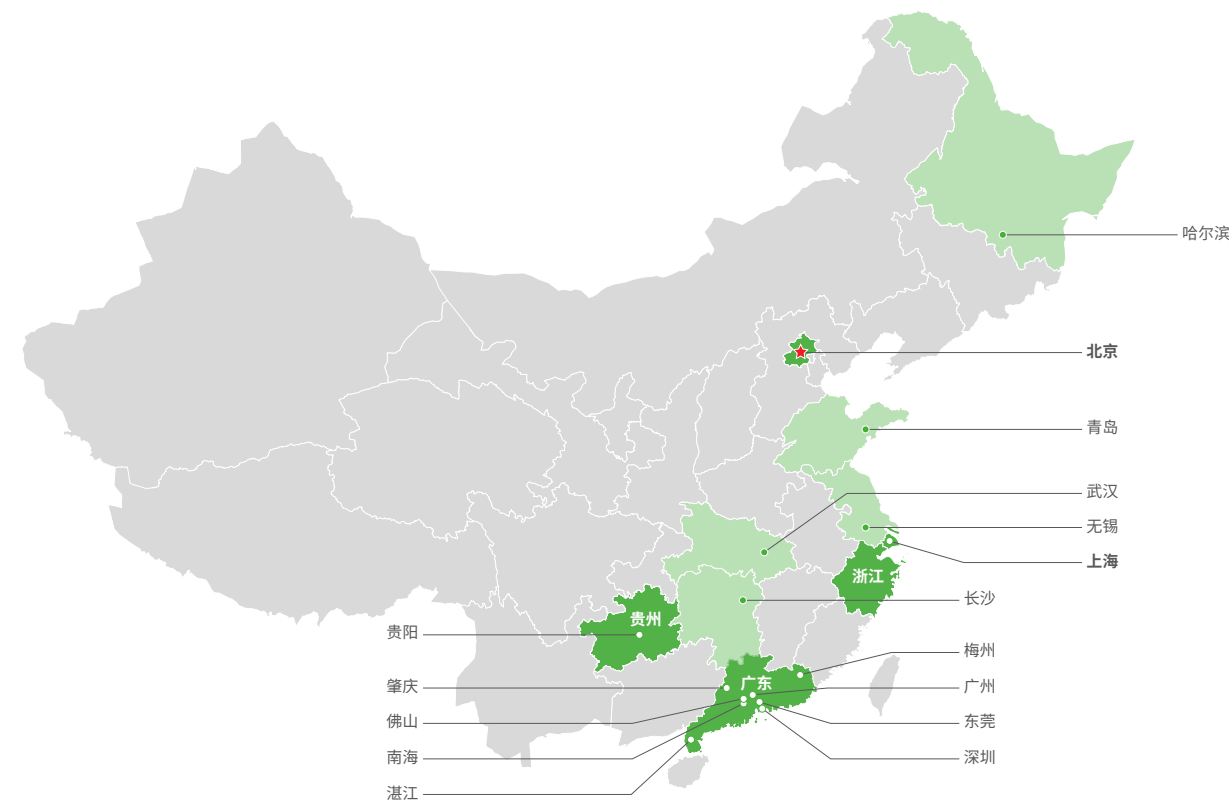


图 2 各地平台地理分布

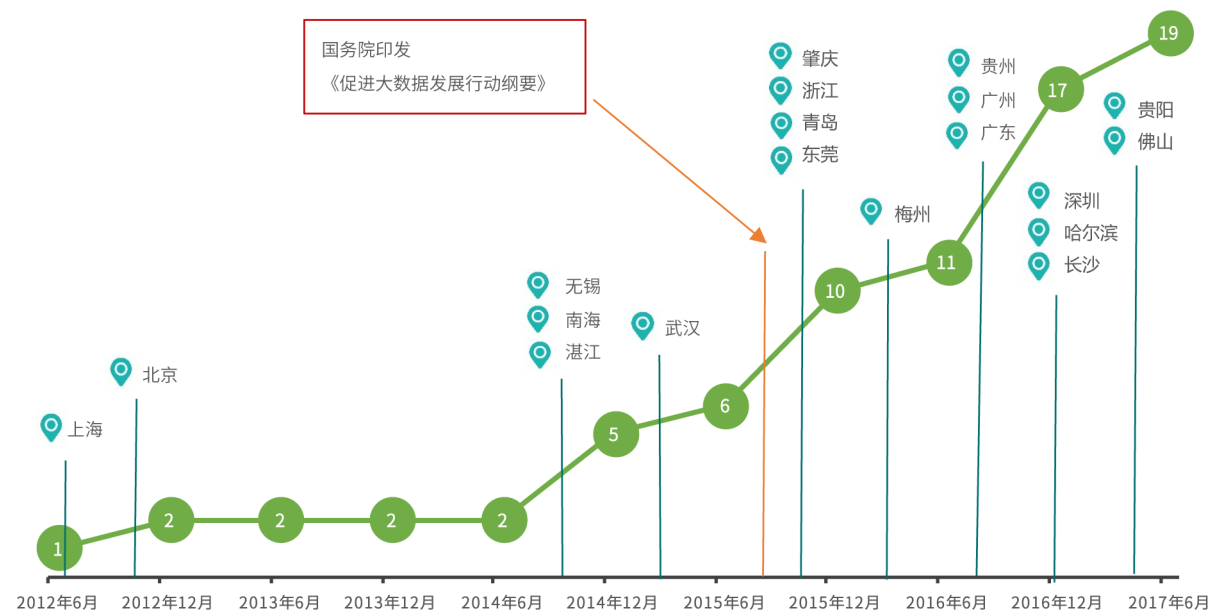


图 1 各地平台上线时间

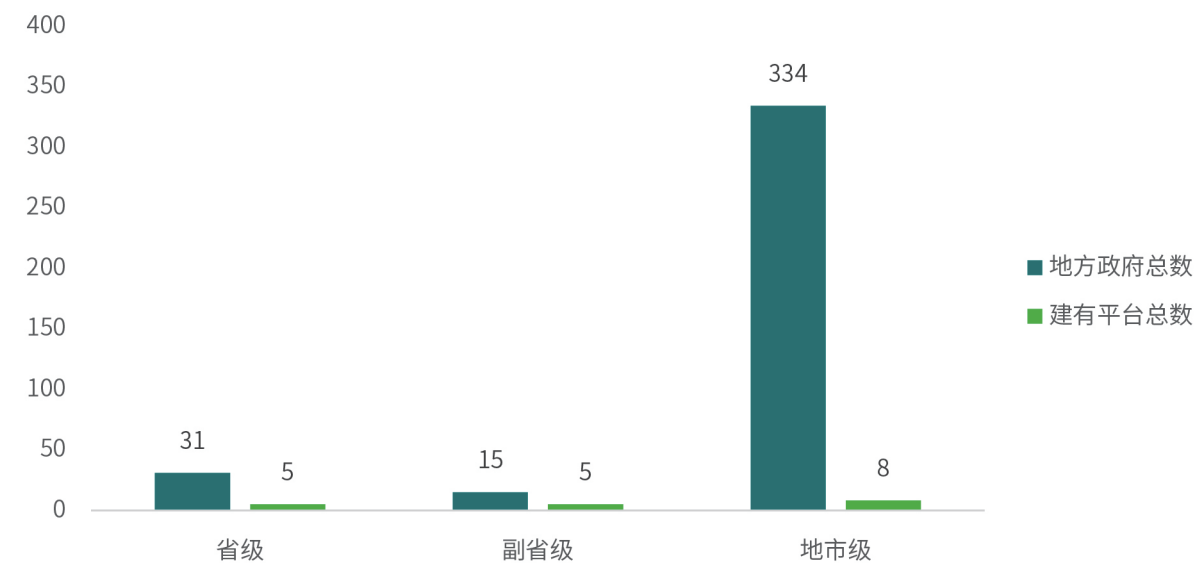


图 3 各地平台行政层级分布

(3) 行政层级

报告对各地平台所属的地方政府行政层级及其所占比

例进行了分析（如图 3）。目前各地上线的平台中地级市数量最多，省级行政区和副省级城市数量相当，然而我国共有 31 个省级行政区（港澳台除外）、15 个副省级城市和 334

3. 主要发现

Key Findings

个地级市，与我国省级行政区、副省级城市、地级行政区的总数相比，已开放数据的地方所占比例普遍偏低。

(4) 平台类型

从各地开放数据平台的类型来看，绝大多数平台以统一专有方式呈现，即开放数据统一汇聚在一个专门的平台上进行开放。只有梅州和肇庆为统一嵌入式，即开放数据统一汇聚为一个栏目板块，嵌入在政府门户网站或政务服务网站上。统一专有式平台更有利于用户通过一站集中的方式发现、获取和利用数据，也是目前国际上普遍采用的方式。

3.2 开放数据集总量

数据集，是一种由数据所组成的集合，通常以表格形式出现，“列”代表一个特定变量，“行”对应于一个样本单位。政府数据开放平台通常以下载或 API 接口的形式开放数据集。个别平台上出现的以下三类情况，在报告中将不视为有效的开放数据集：（1）数据集名称下不存在可直接下载或通过接口获取的数据集；（2）数据集中仅有 0 行或 1 行数据（多为将一个数据集拆分出的单行数据，或未整合成一个数据集的单行数据）；（3）数据集名称下提供的是网页链接，且链接跳转后出现的是无法通过下载或接口形式获取的文本内容。

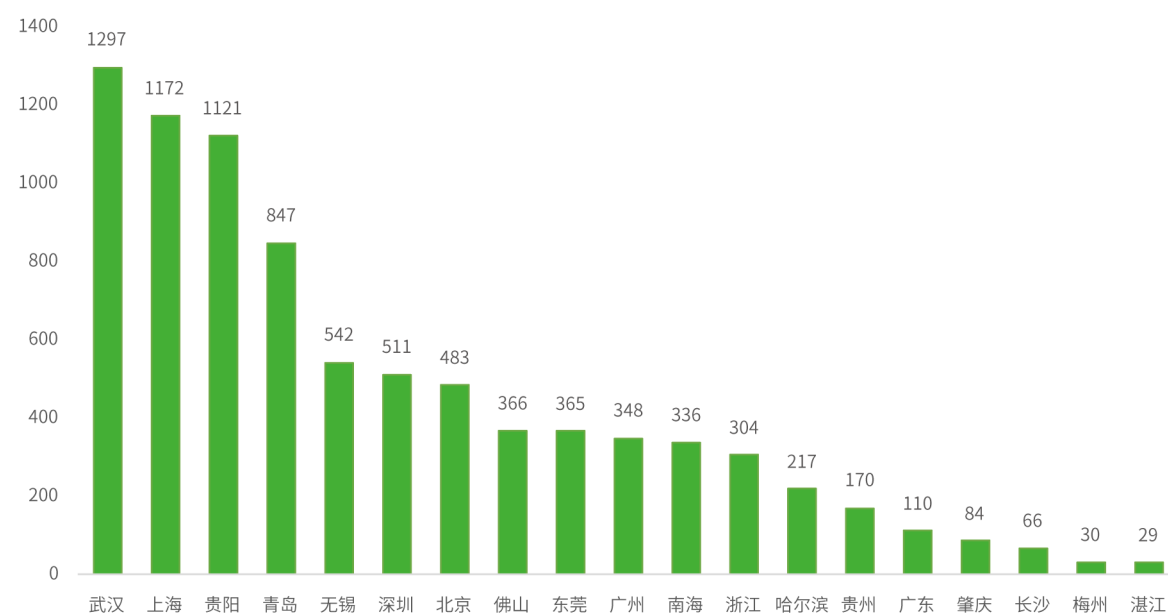


图 4 各地平台上开放的数据集总量

同时，报告将各地平台上出现的以下三类情况只视为一个有效的数据集：（1）同样名称的数据集重复出现；（2）平台上标注的数据集名称不同，但实际下载后的数据集具有相同的名称与内容；（3）同一个数据集按行政区划与部门

等被分成多个数据结构相同的数据集。

各地开放数据平台开放的有效数据集总量（含直接下载和 API 接口开放）如图 4 所示。在 19 个开放数据平台中，

武汉开放的数据集最多，与上海、贵阳处于第一梯队，都开放了超过 1000 个有效数据集或接口；其次是青岛，开放了 847 个有效数据集。其后，无锡、深圳、北京开放了 400 多个的数据集；佛山、东莞、广州、南海、浙江等地开放了 300 多个。

3.3 数据开放度

关于开放数据的格式标准，万维网发明者 Tim Berners-Lee 提出了一个开放数据五星标准⁷：一星：基于开放授权在网络上开放数据，对数据格式不做要求，可能采用 PDF、JPEG 等格式；二星：以可机读、结构化格式开放数据（如采用 EXCEL 格式而非扫描格式）；三星：在满足二星标准的基础上，以非专属开放格式开放数据（如采用 CSV 格式代替 EXCEL 格式）；四星：在满足以上要求的基础上，为每一个数据集设置固定的 URL 链接，便于使用者

发现和链接到数据集的具体位置；五星：在满足以上要求的基础上，数据之间实现关联，以展现数据的背景。本报告基于该标准和其他有关开放数据格式的标准，重点对各地开放平台上的数据集的开放度进行评估。

3.3.1 技术性开放

(1) 可机读格式

为方便用户获取和利用数据，数据集应以可机读格式开放，该格式能被计算机自动读取与处理，如 XLS、CSV、JSON、XML 等格式。图 5 展示了各地平台上可下载数据集总量与可机读数据集总量的对比。总体来看，各地平台上满足可机读格式的数据集比例较高，贵阳、上海、青岛、东莞、佛山、南海、广州、哈尔滨等平台上开放的可下载数据集基本都满足了可机读格式。但也有一些地方平台虽然开放的数

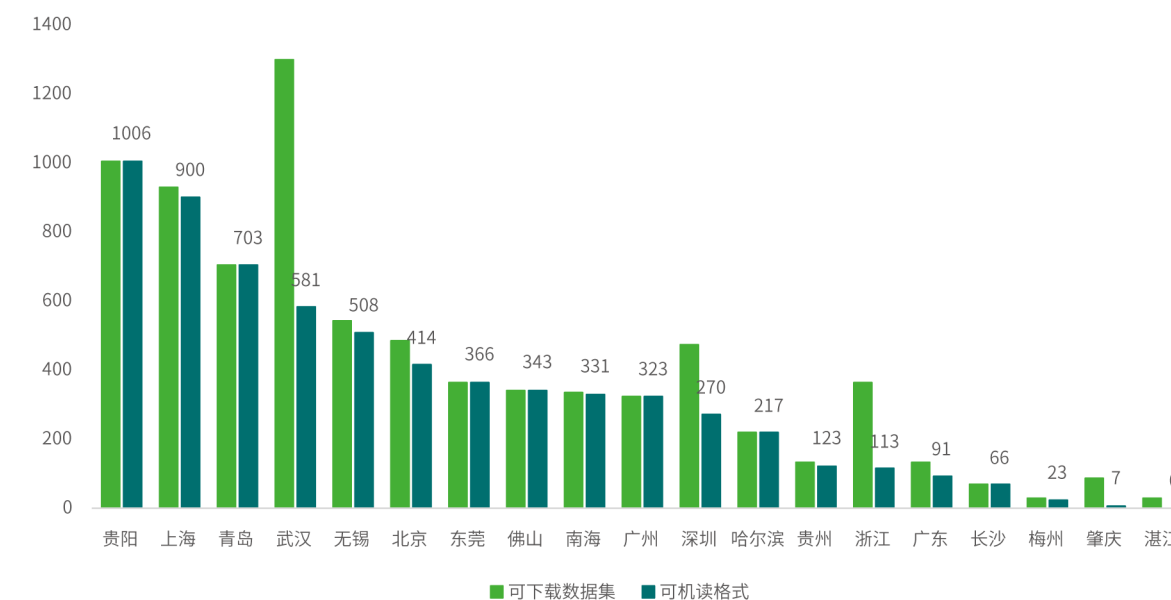


图 5 各地平台可下载数据集与可机读格式数据集数量

7 <https://www.w3.org/DesignIssues/LinkedData.html>

3. 主要发现 Key Findings

据集总量不小，但其中不少数据集为 DOC、PDF、JPG 等不可机读格式。还有部分平台虽然提供了可机读格式的数据集，但内容多为加工归总后的统计数据，并不满足开放完整的、原始的、可再利用的数据集的开放数据原则。

(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）应用程序才能访问。例如 CSV 是开放格式，而 XLS 则不是。

图 6 是各地方平台上可下载数据集总量与开放格式总量的对比，贵阳平台上开放格式的数据集数量最多，所有的数据集都可通过开放格式下载。东莞、佛山、南海、广州、哈尔滨、长沙等地方平台上提供的数据集也都满足开放格式

的标准。北京平台上开放的大多数数据集也达到了开放格式标准。此外，约有一半的地方平台没有提供任何开放格式的数据集，而多以 XLS 格式为主。

(3) API 接口

除了数据直接下载以外，有些地方平台还提供了应用程序编程接口（API 接口，Application Programming Interface）。与通过直接下载方式获取数据集相比，接口方式使用用户可通过参数实时高效地获取所需数据，尤其适用于开发应用程序的需求。深圳、北京、东莞三个地方平台为每个数据集提供了接口（见图 7）。其中，深圳开通的接口数量最多。此外，佛山、上海、贵阳等地也提供了一定数量的接口。然而，仍有将近一半的平台没有或很少提供 API 接口，还有个别地方平台上的 API 接口多数处于无法被调用的状态。

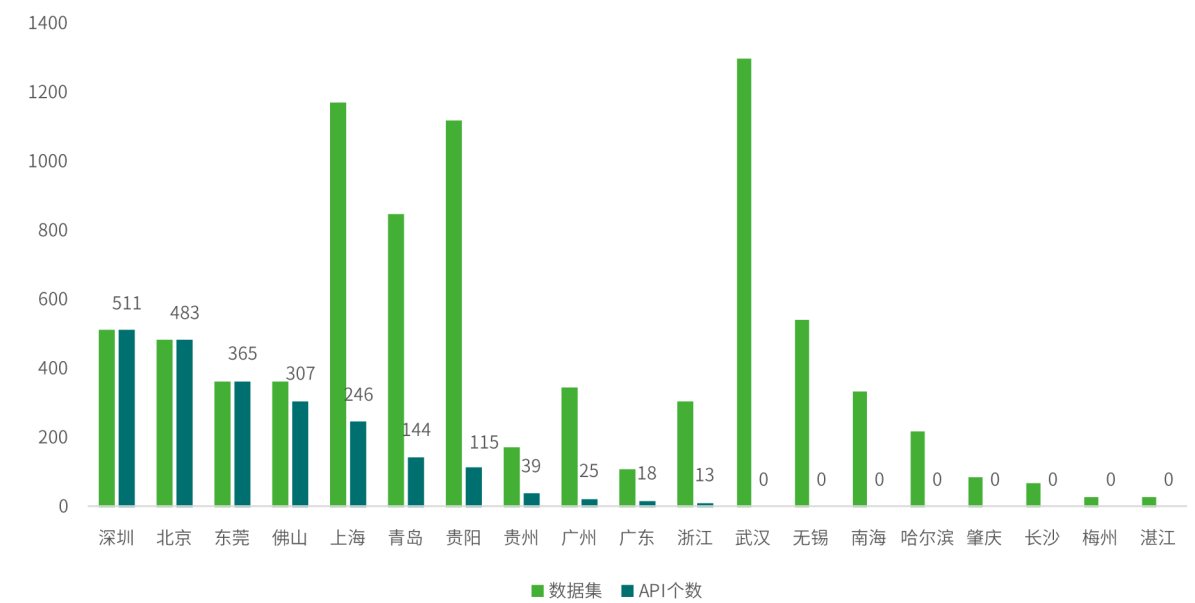


图 7 各地平台数据集总量与 API 接口数量

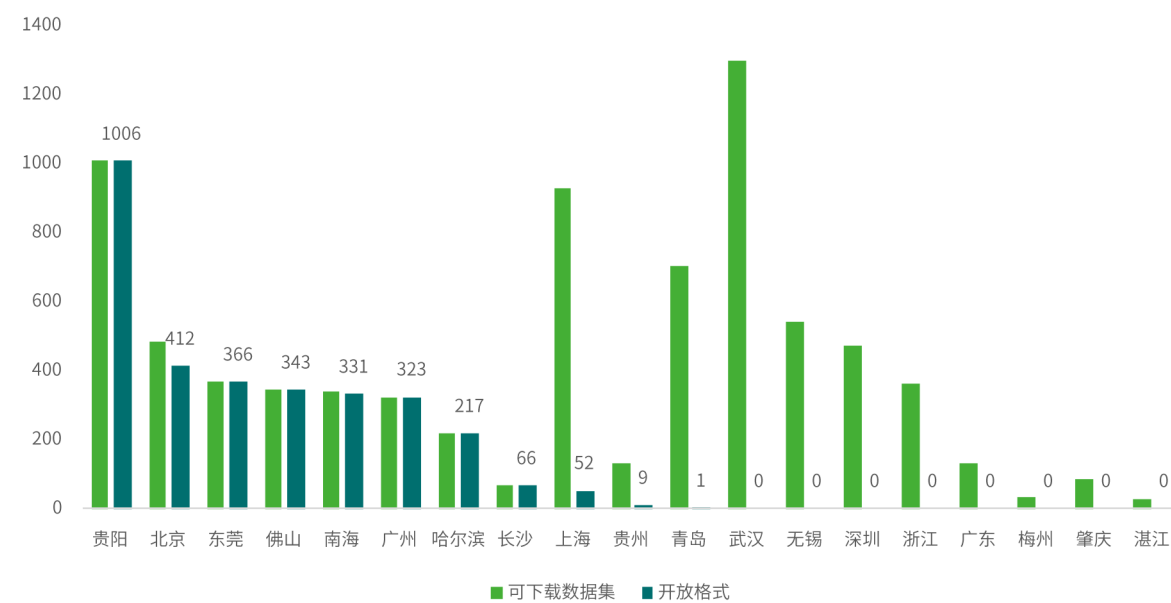


图 6 各地平台可下载数据集与开放格式数据集数量

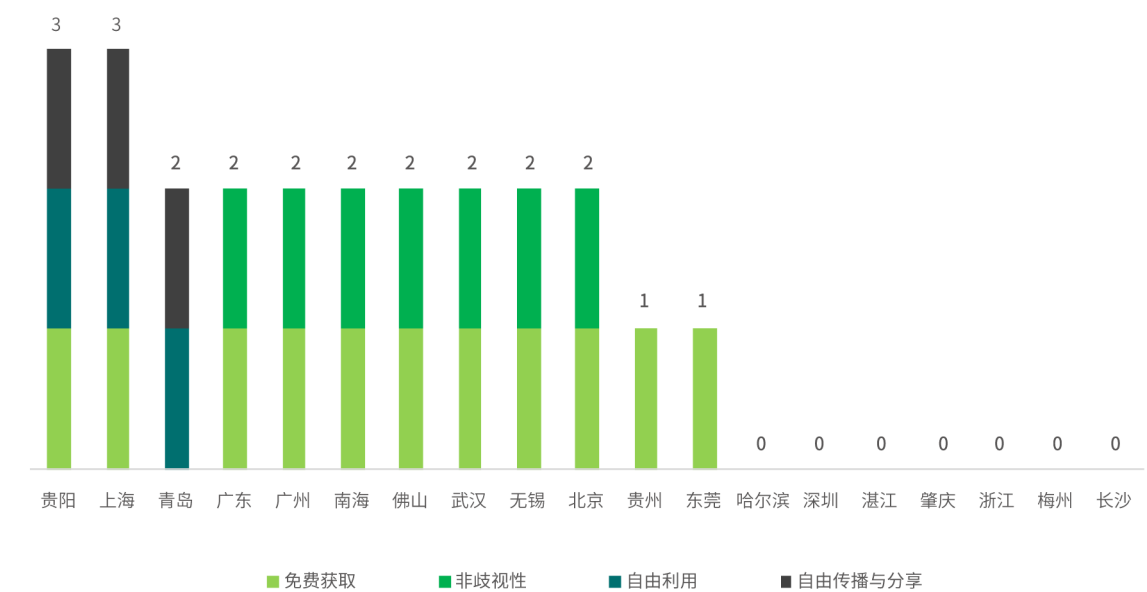


图 8 各地平台开放授权详情

3. 主要发现

Key Findings

3.3.2 开放授权

开放数据应通过数据开放授权协议从法律上保障数据的开放性，数据开放授权通常包含在开放数据平台的网站声明、免责条款或服务协议中。报告发现，在 19 个政府数据开放平台中，共有 12 个平台配有数据开放授权。7 个平台缺少专门的数据开放授权，占评估总体的 36.8%，这其中有 4 个平台的网站声明与政府门户网站共享，并非专门针对开放数据平台制定，也没有专门涉及数据开放授权的内容。

开放授权协议的内容应明确授予用户免费获取、不受歧视、自由利用、自由传播与分享“开放数据”的权利。目前没有一个地方平台的授权协议全部明确授予了上述四项权利。上海和贵阳满足了其中三项指标；青岛、广东、广州、佛山南海、佛山、武汉、无锡和北京满足了其中两项；贵州和东莞满足了一项；其余地方在四项指标上未明确提及，或语焉不详（见图 8）。

(1) 免费获取

免费获取是指平台在开放授权中明确授予用户免费获取和利用“开放数据”的权利。政府数据作为公共资源，原则上应免费向社会开放，除非需要对数据进行额外的增值加工和针对少数人的个性化加工等。目前各地平台上的相应条款分为“免费且未设时限”、“现阶段免费”、“未明确是否免费”三类。其中，贵阳、东莞的条款中明确说明数据免费且未设时限，用户可永久无偿获取数据平台所提供的所有数据资源。其他大部分地方平台则在服务协议中提到“现阶段免费”，但设置有模糊的期限或限制，如“保留收费权利”等表述；其他少数城市未明确表述是否免费（见图 9）。

(2) 非歧视性

开放授权是指平台明确授予任何用户平等访问、获取、

使用和分享“开放数据”的权利。目前各地平台上的相应条款分为“明确保障非歧视性”、“未明确保障非歧视性”两类情况。多数地方的开放数据平台明确保障了数据开放的非歧视性，对任何用户都予以平等的获取和利用权限；但上海、贵阳、贵州、东莞、青岛等地未明确提及是否保障非歧视性（见图 10）。

(3) 自由利用

开放授权应明确授予用户不受限制地对“开放数据”进

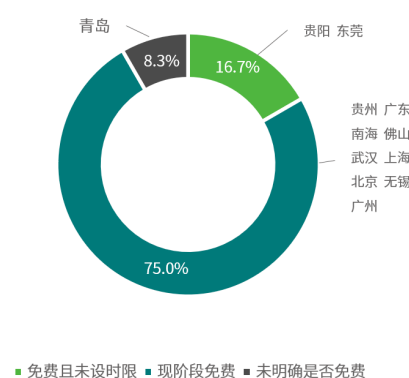


图 9 开放数据免费获取比例

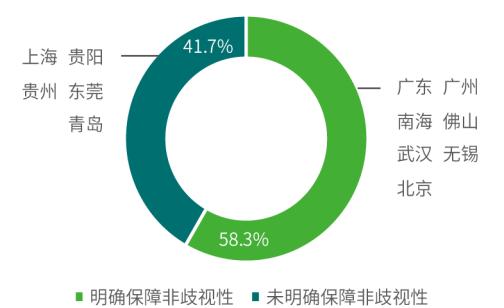


图 10 开放数据非歧视性比例

行商业和非商业性利用的权利。目前各地平台上相应条款分为“明确允许自由利用”和“未明确提及可自由利用”两类。其中上海、青岛和贵阳明确表明用户可“进行商业或非商业性的开发活动”、“享有增值利用的权利”或“拥有自由利用的权利”，其余 9 个地方均未对用户利用数据的权利做出明确说明（见图 11）。

(4) 自由传播与分享

开放授权应明确授予用户可自由传播和分享“开放数据”的权利。目前各地平台上的相应条款分为“可自由传播”、“自由传播受限”、“未提及可自由传播”三类情况。贵阳、青岛、上海授予用户享有免费传播现有开放数据的权利；为保障用户免费获取数据的权利，青岛特别说明了用户不得有偿转让获取的数据。其他大部分地方在使用条款中要求“用户不得有偿或无偿转让数据资源”，限制有偿转让数据有利于保障用户免费获取数据的权利，但限制用户无偿传播和分享已开放的数据并不符合开放数据的原则，在实际操作中也很难发现和限制。其他地方未对用户自由传播与分享的权限做出明确说明（见图 12）。

3.3.3 元数据提供

政府数据开放平台为开放数据集提供元数据有助于数据利用者清楚地了解和理解数据集的内容与背景，从而更好地获取和利用数据。本报告对各地平台上的可下载数据集所提供的元数据进行了分析。

(1) 平台元数据条目

图 13 显示了各地平台上提供的元数据条目数量。贵阳提供的元数据条目数最多，其次是哈尔滨与广州，也有少数平台只提供了较少的元数据条目。

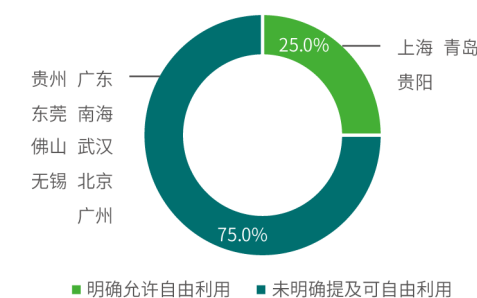


图 11 开放数据自由利用比例

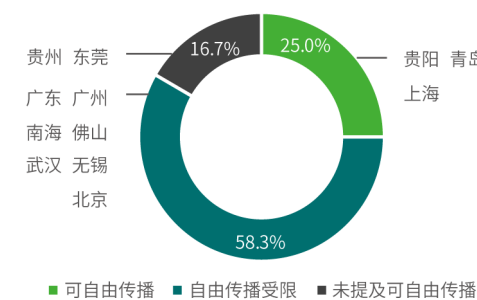


图 12 开放数据自由传播与分享比例

(2) 基本元数据覆盖率

综合梳理国际开放数据平台上提供的基本元数据条目以及目前我国半数以上的平台已实际提供的元数据条目，报告确定了以下 13 个条目作为开放数据集基本的元数据条目，包括数据名称、摘要、关键字、数据主题、下载格式、数据提供方、发布日期、更新日期、更新频率、访问量、数据项（字段）、数据量等。其中数据量是指数据集所含的行数（或行数乘以列数），数据项是指数据集中每一列的具体名称，两者有助于利用者更好地了解数据结构与内容。

3. 主要发现 Key Findings

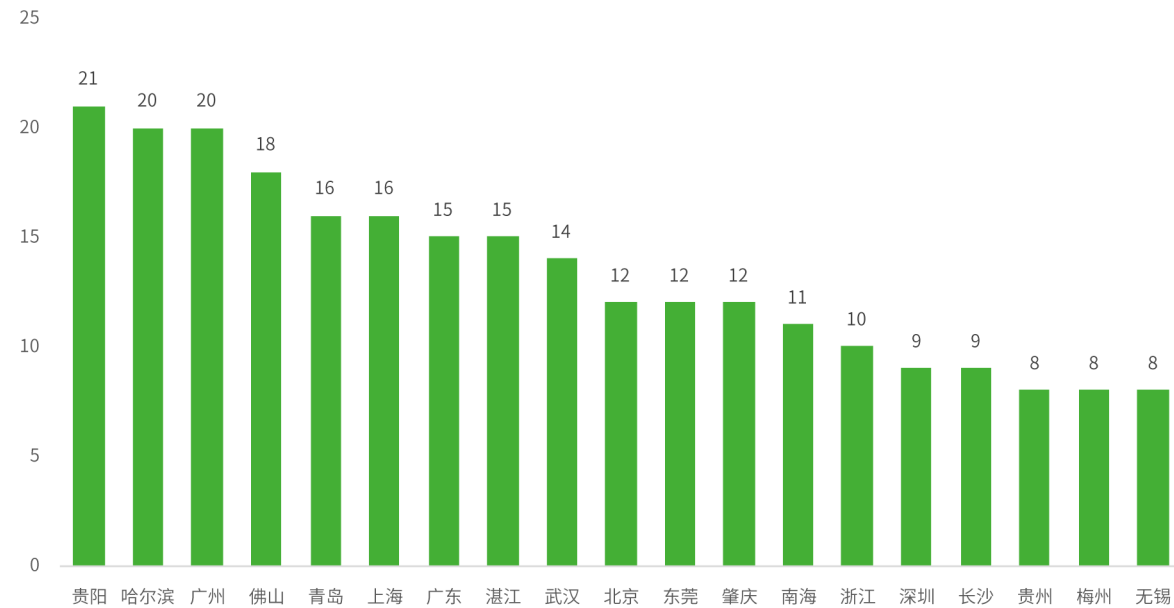


图 13 各地平台元数据条目

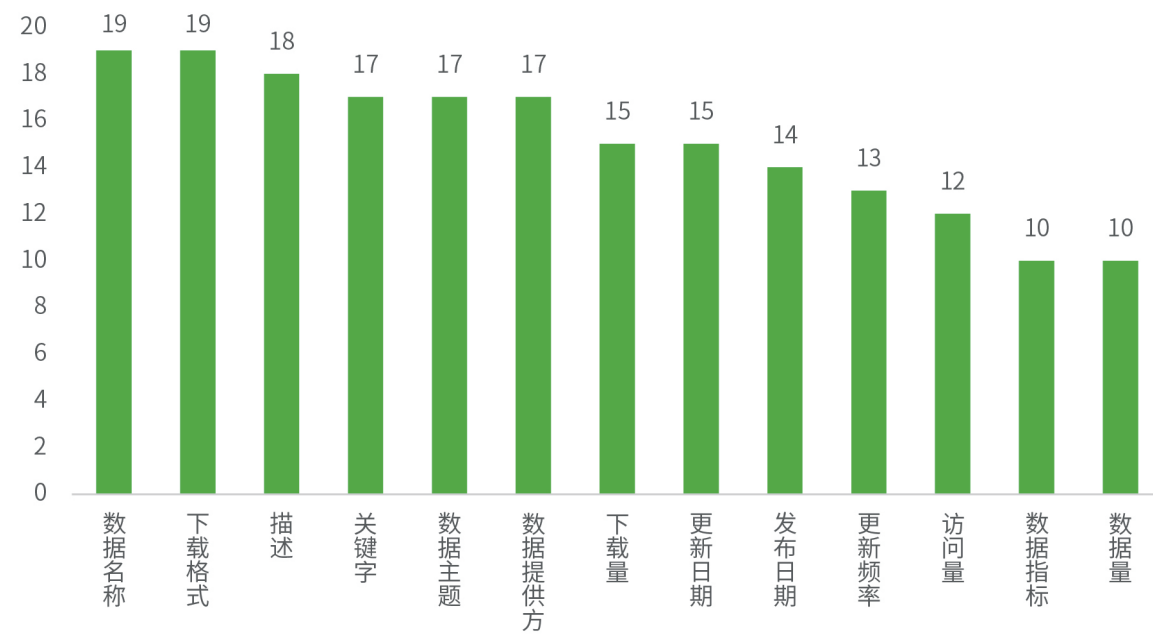


图 14 基本元数据条目在各地平台的分布数

图 14 表示以上 13 个基本元数据条目在 19 个地方平台的分布情况。目前，所有 19 个地方平台都已提供了数据集名称与下载格式，大多数平台已提供数据描述、数据主题、

关键字（标签）与数据提供方。然而，目前能提供更新频率、访问量、数据项、数据量条目的地方平台还相对较少。

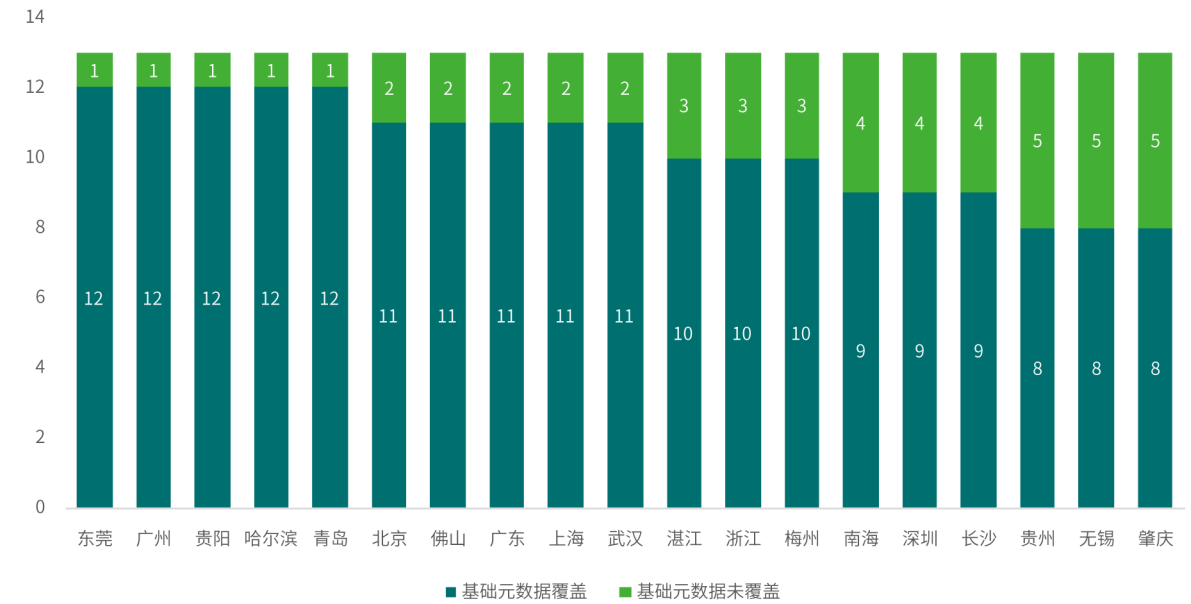


图 15 各地平台基本元数据条目覆盖数

	数据名称	描述	关键字	数据主题	下载格式	数据提供方	下载量	更新日期	发布日期	更新频率	访问量	数据指标	数据量
东莞	●	●	●		●	●	●	●	●	●	●	●	●
广州	●	●	●	●	●	●	●	●	●	●	●	●	●
贵阳	●	●	●	●	●	●	●	●	●	●	●	●	●
哈尔滨	●	●	●	●	●	●	●	●	●	●	●	●	●
青岛	●		●	●	●	●	●	●	●	●	●	●	●
北京	●	●	●	●	●	●	●	●	●	●	●	●	●
佛山	●	●	●	●	●	●	●	●	●	●	●	●	●
广东	●	●	●	●	●	●	●	●	●	●	●	●	●
上海	●	●	●	●	●	●	●	●	●	●	●	●	●
武汉	●	●	●	●	●	●	●	●	●	●	●	●	●
湛江	●	●	●		●	●	●	●	●	●	●		●
浙江	●	●	●	●	●	●	●	●	●	●			
梅州	●	●		●	●	●	●	●	●			●	
南海	●	●	●		●	●	●	●	●			●	
深圳	●	●		●	●	●	●	●	●			●	
长沙	●	●		●	●	●	●	●	●		●	●	●
贵州	●	●	●	●	●	●		●		●	●		
无锡	●	●	●	●	●	●			●	●	●		
肇庆	●	●	●	●	●	●			●	●			

图 16 各地平台基本元数据条目详情列表

图 15 表示各地平台提供的基本元数据条目的数量，目前还没有一个平台提供了所有的基本元数据，东莞、广州、贵阳、哈尔滨与青岛覆盖了 12 个基础元数据条目，仅缺少

一条；其次是北京、上海、广东等地，都缺少两条。各地平台提供的基本元数据条目详情可参见图16，无圆点的空格即为该地方缺少的元数据条目。

3. 主要发现

Key Findings

3.4 数据覆盖面

3.4.1 主题覆盖率

开放各个主题领域的数据集有利于提高数据开放的广度和覆盖面，有利于数据利用者充分获取和融合来自多种领域的的数据，进行深度的挖掘利用。2015年8月国务院印发的《促进大数据发展行动纲要》提出要“优先推动信用、交通、医疗、卫生、就业、社保、地理、文化、教育、科技、资源、农业、环境、安监、金融、质量、统计、气象、海洋、

企业登记监管等民生保障服务相关领域的政府数据集向社会开放”。基于行动纲要的要求，结合国内外开放数据平台上实际开放的重点数据集主题，报告将开放数据主题归纳为财税金融、经贸物流、交通出行、机构团体、文化休闲、医疗卫生、教育科技、社会民生、资源环境、城建住房、公共安全、农业农村、社保就业、信用服务共14个大类。图17体现了各地平台在14个主题下所开放的数据集个数。其中，经贸物流的数据集开放数量最高，包含了经济建设和企业服务等内容；其次为社会民生、机构团体的数据集，数据开放相对较少的主题为财税金融、农业农村和信用服务等。

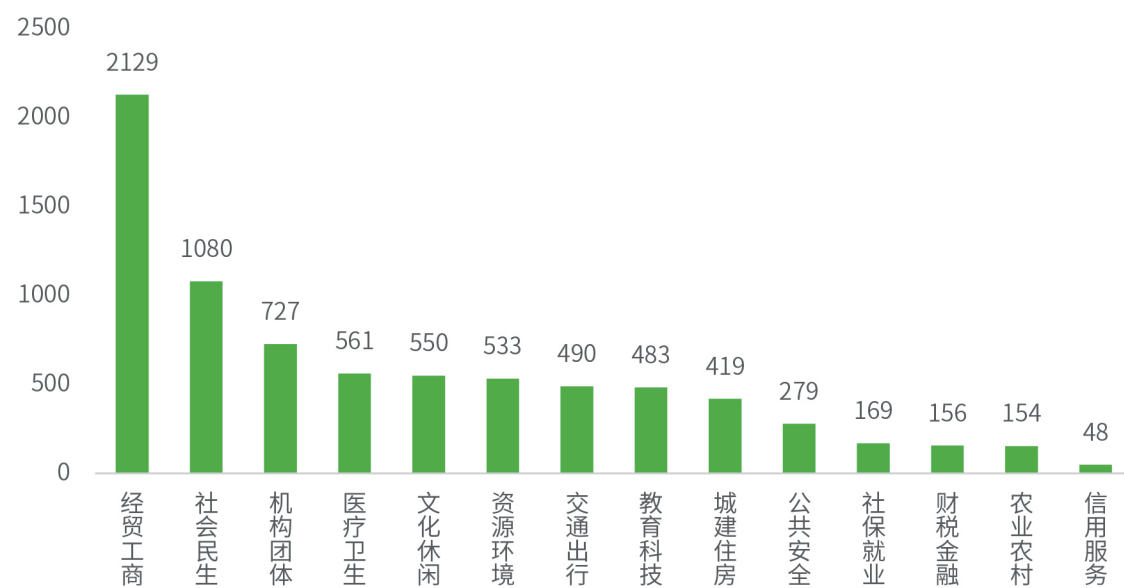


图 17 各主题包含的数据集个数

报告进一步分析了各地方平台开放数据集的主题领域覆盖面，图18显示了各地已覆盖的主题个数。在14个主题领域中，不同地区间的主题分布情况差异显著。广州、贵阳、青岛开放的数据集主题覆盖面最高，只有一个未覆盖主题，其次为北京、南海、无锡等地。开放数据集总量较少的

地方其主题覆盖率也相应较低。这些地方之间的差异可能与各地发展重点、开放意愿和数据储备基础等因素有关。

某些地方平台上主题领域的用词与社会上的数据利用者或普通公众的用词习惯有差异，过于宏大和宽泛，更多反

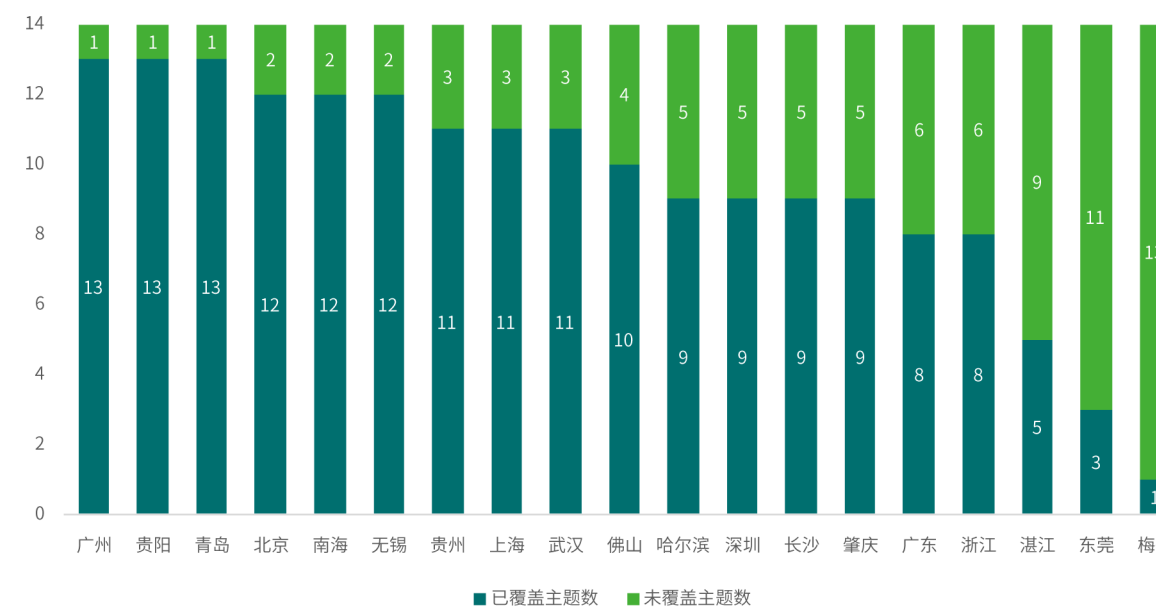


图 18 各地平台主题覆盖数

映了政府部门的视角，如“经济建设”和“社会发展”等，还有些地方的主题领域用词则过于模糊，如“综合信息”和“工作结果”，不利于用户发现、理解和获取其需要的数据。

3.4.2 部门覆盖率

开放数据集的部门覆盖率反映了一个地方政府的各个部门对于数据开放工作的参与程度，也体现了数据集的丰富程度。报告首先梳理了超过半数的地方平台开放数据集所覆盖的部门作为“主要数据提供部门”，由于不同地方具有相同职能的部门在名称上存在差异，报告对这些部门的名称做了相应的合并与调整。

图19为各地平台上主要数据提供部门所提供的数据集数量。由图可见，各地平台上来自统计局的数据集数量最多，远高于其后的商务、交通、卫生计生、公安、教育、经信等

部门，而提供数据集数量相对较少的是财政、气象和档案等部门。图中的不同色条的纵向宽度反映了各地平台上不同数据提供部门的数据集数量，如上海提供最多数据的部门为交通、卫生计生与公安等，统计部门数据所占相对较少。统计部门提供的数据多为经过归总加工后的宏观数据，各地还应继续提升各个具体业务部门的开放数据集比例。

同时，报告还分析了各地平台数据提供部门数量占该地方政府职能部门总数（数据来自各地政府门户网站）的比例，以评估该地方的开放数据部门的覆盖面。结果如图20，贵阳的部门覆盖率超过了9成，此后是上海与南海，均超过了7成，仍有部分地方平台的部门覆盖率不到一半，尚有较大提升空间。

3. 主要发现 Key Findings

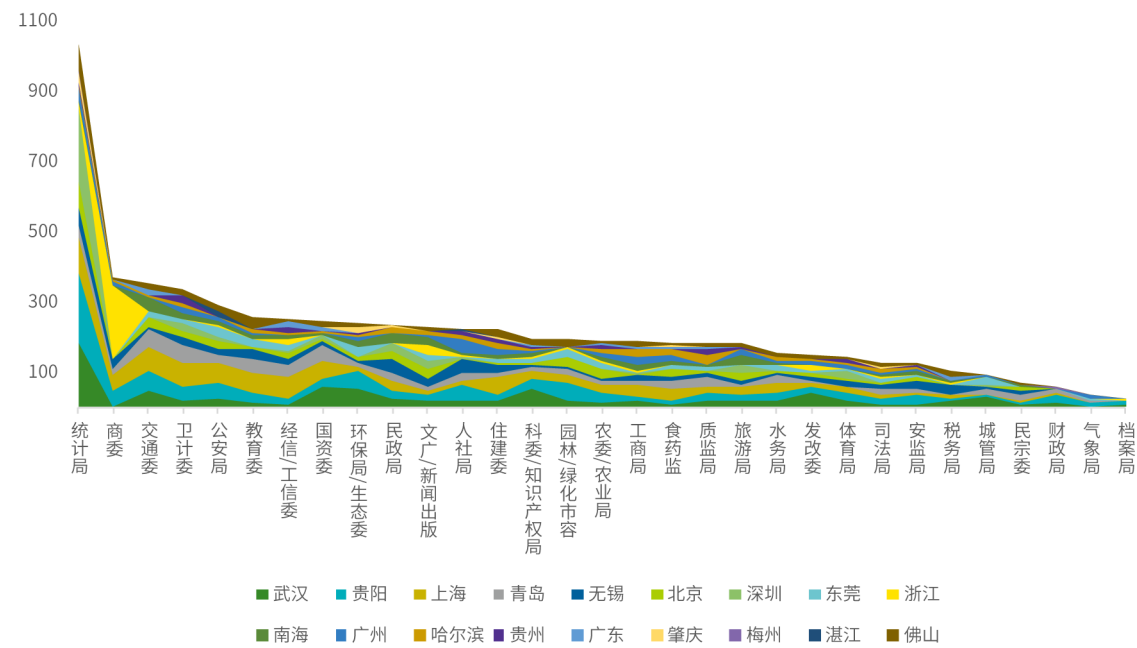


图 19 各地平台主要数据提供部门数据集数量分布

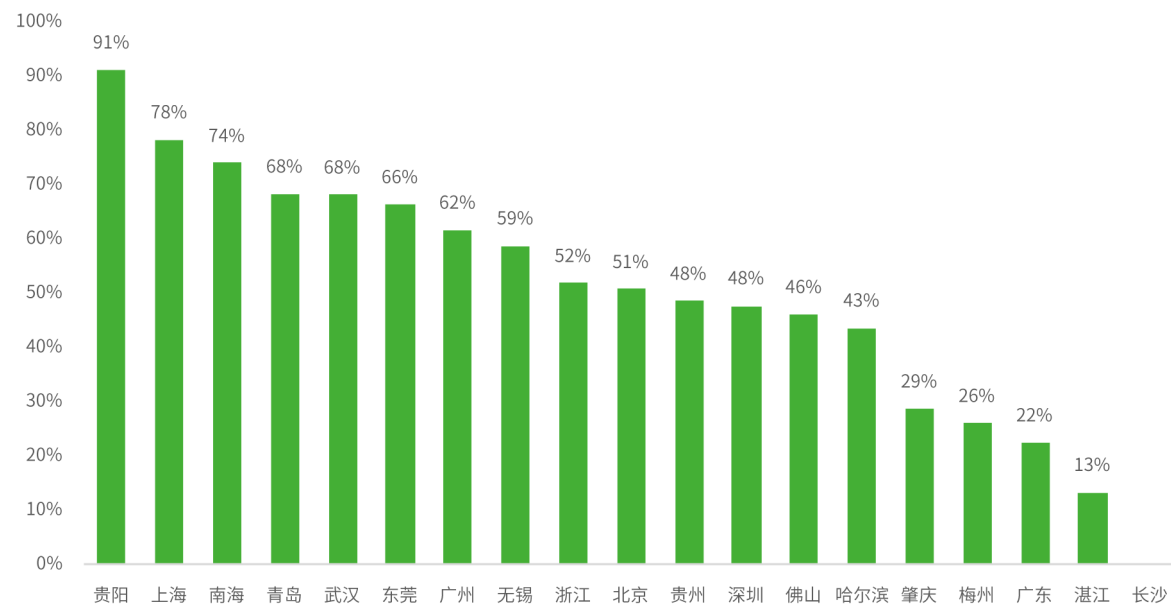


图 20 数据提供部门占部门总数的覆盖率



图 22 各地开放数据集名称中出现频次最高的 48 个描述性限定词

3.4.3 关键词覆盖率

报告对各地平台上数据集名称进行了文本分析，发现了一批出现频次较高的关键词，这些关键词反映了各地开放数据的内容及其分布。如图 21 所示，各地开放数据集中出现频次最高的关键词是“信息”和“情况”，远远高于其后的“企业”、“机构”、“统计”等。这类词汇仍旧体现了政府部门的用词习惯，而不是社会公众更容易理解的词汇。值得注意的是，作为政府数据开放平台，“数据”一词出现的频次却远小于其它名词，也反映了政府部门的惯性思维。

图 22 为各地数据集中高频出现的描述性限定词，这些限定词描述了数据集的主要内容和方向。由图可见，“经营”、“生产”和“预算”出现的频次最高，其后“审批”、“许



图 21 各地开放数据集名称中出现频次最高的 33 个名词

3. 主要发现

Key Findings

可”、“机动车”、“企业名单”、“蔬菜”等词也有较高的出现频次。在以上高频描述性限定词与名词的搭配中，“审批信息”、“公开信息”、“许可机构”、“经营企业名录”、“预算公开”等数据集出现频次较高，这些数据集的来源多为政府信息公开的内容，以文本为主，数字较少，更多为统计加工后的数据，具有高价值的、原始的、完整的数据集较少。

报告将各地开放的数据集名称中高频次出现的 48 个描述性限定词确定为“基本关键词”，然后分析各地平台数据

集名称对于这些基本关键词的覆盖率。如图 23 所示，覆盖率最高的是贵阳，达到了全部覆盖，其次是无锡、武汉、青岛和上海，基本关键词覆盖率都超过了 9 成，少数城市覆盖率较低，不到两成。

各地开放的数据集名称对于基础关键词的覆盖详情如图 24 所示，数字所在单元格颜色越深，代表涉及到该关键词的数据集越多，颜色越浅则代表该项数据集越少。

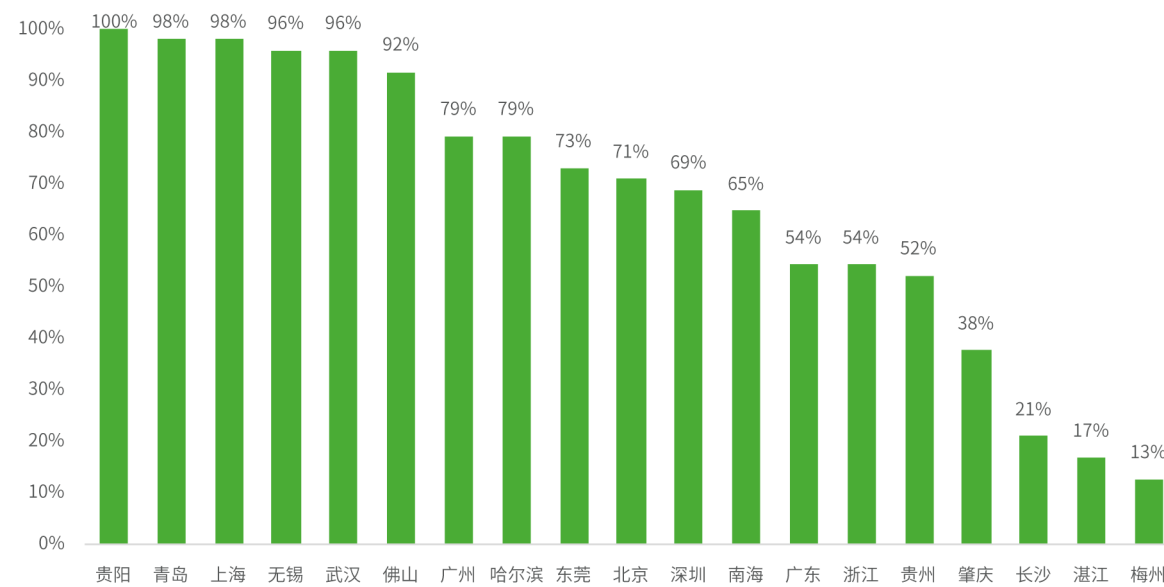


图 23 各地开放数据集名称基本关键词覆盖率

	北京	东莞	佛山	广东	广州	贵阳	贵州	哈尔滨	梅州	南海	青岛	上海	深圳	无锡	武汉	湛江	长沙	肇庆	浙江
经营	15	34	8	4	9	38	1	1	2	12	24	2	33	19	24	0	0	1	14
生产	43	18	12	7	6	41	0	7	0	11	21	45	15	23	24	1	2	1	19
预算	0	0	1	0	0	41	2	0	0	0	6	6	2	1	111	0	0	0	0
许可	21	9	17	38	54	58	0	11	1	14	21	5	72	18	21	5	0	1	24
审批	2	31	2	6	4	1	0	4	1	1	3	48	5	8	16	0	0	0	0
投资	0	3	5	1	1	22	1	3	0	4	6	18	32	18	23	0	0	0	6
机动车	1	3	4	3	3	3	0	2	1	9	13	15	0	3	8	1	1	1	0
建设	4	15	12	9	7	37	4	9	0	3	14	28	14	13	29	0	0	0	26
蔬菜	0	0	1	0	0	1	0	0	0	0	1	8	0	0	0	0	0	0	86
旅游	9	1	5	0	9	19	3	0	0	0	7	14	5	6	13	0	0	2	0
企业名单	7	3	6	3	9	7	0	1	0	2	8	4	0	14	13	4	0	4	11
价格	3	0	11	0	1	4	3	0	0	0	19	14	4	12	33	0	0	0	152
财政	0	0	1	0	1	2	2	0	0	0	4	1	3	5	37	0	0	1	2
收费	5	0	0	1	5	12	1	1	0	0	17	6	0	12	23	0	0	0	0
市场	2	4	3	2	19	12	2	2	0	2	8	5	3	5	18	0	1	1	64
学校	6	7	19	0	4	8	0	6	0	4	11	12	9	5	15	0	0	1	2
卫生	8	4	4	0	5	24	14	5	1	1	5	13	0	12	16	0	0	3	1
建设项目	1	6	2	3	4	1	0	4	0	1	4	7	7	1	7	0	0	0	13
规划	2	4	2	1	0	17	0	3	0	1	4	21	3	3	27	0	0	0	0
收支	0	0	1	0	0	8	1	0	0	0	5	3	3	4	51	0	0	0	0
违法	0	16	4	5	5	7	0	2	0	1	4	5	33	0	7	0	0	0	0
生产总值	16	0	4	0	2	5	0	1	0	0	3	17	6	4	4	1	0	1	0
行政处罚	1	15	5	3	18	4	0	2	1	1	4	7	7	1	0	0	0	0	0
农业	9	5	6	2	3	5	2	2	0	3	4	8	6	3	12	0	1	1	2
药品	7	7	0	1	4	14	1	2	0	0	6	9	0	1	5	0	0	3	2
道路	1	1	1	3	0	19	0	2	0	2	11	12	0	6	1	0	0	0	1
监测	2	1	2	0	3	13	3	0	0	2	11	11	0	5	26	0	0	0	0
培训	3	6	2	1	4	11	0	5	0	6	3	8	0	1	5	3	0	0	1
受理	0	0	2	0	4	8	0	1	0	2	4	18	2	4	5	0	0	0	0
固定资产	0	0	2	0	0	11	0	1	0	1	2	7	18	3	11	0	0	0	1
房地产	0	1	9	0	3	6	0	6	0	0	4	9	9	3	12	1	1	0	0
医疗机构	4	2	2	0	5	3	2	1	0	4	8	15	4	1	6	0	0	0	0
资格	4	11	2	2	2	13	0	1	0	6	11	17	0	7	12	0	0	0	2
食品	8	4	3	1	4	1	1	1	0	5	8	16	0	4	9	0	2	1	0
资金	0	1	3	0	1	3	1	0	0	0	1	17	2	7	56	0	0	0	0
教育	5	5	13	1	3	19	2	1	0	4	17	17	0	15	13	0	1	0	0
销售	3	2	2	2	0	4	3	2	0	2	5	5	7	7	11	0	0	1	6
用地	0	3	2	6	0	7	0	2	0	0	1	9	0	6	16	0	0	0	0
设施	3	11	2	1	2	11	3	2	0	0	4	22	5	3	13	0	1	1	1
零售	2	16	2	1	6	17	0	2	0	3	5	17	7	3	11	0	1	1	4
生活	1	2	3	0	1	6	0	7	0	0	4	8	3	2	18	0	0	0	4
科技	0	5	1	1	3	12	8	2	0	1	2	4	1	2	11	0	1	2	1
医院	12	0	1	0	1	8	3	2	0	2	7	1	5	4	5	2	0	0	1
农村	2	0	2	0	8	21	1	4	0	0	3	4	0	3	18	0	0	0	0
从业人员	3	7	3	0	0	9	0	3	0	0	3	4	4	4	5	0	0	0	0
住房	0	0	4	0	5	11	1	4	0	2	7	3	0	1	7	0	0	0	1
人力资源	0	1	0	0	22	4	0	0	0	1	2	2	4	2	8	0	0	0	0
经济指标	1	0	0	0	0	9	1	0	0	0	0	0	7	6	6	0	0	0	0

图 24 各地开放数据集名称基本关键词覆盖详情

3. 主要发现

Key Findings

3.5 数据持续性

3.5.1 持续增长

运维政府数据开放平台是一项常态化工作，平台上线后还应持续新增数据集，以不断满足社会对开放数据日益增长的需求。报告根据各地平台上数据集的创建日期来判断该平台数据集是否持续增长，以半年为时段进行跟踪分析，持续增长分值取决于该平台数据集保持持续增长的时段的数量。图 25 显示的是以半年为时段的各地平台数据集持续增长情况，颜色区域表示该时段有新增数据集出现，空白区域表示该时段无新增数据集。

上海平台上线时间最早，2012 年上半年上线，至今数据已保持了 9 个时段的持续增长；北京平台于 2012 年下半年上线，持续增长时段为 6 个。在 2014 年下半年，佛山南海、湛江和无锡的平台上线，但湛江自上线之后便不再新增数据集。2015 年以后新上线的平台基本都能保持数据集增长。

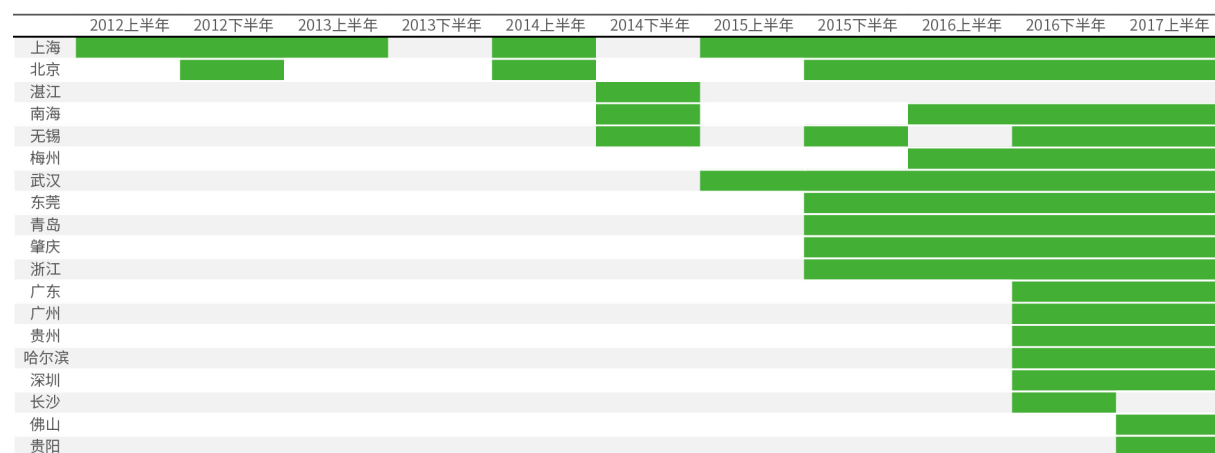


图 25 各地平台上线时间与数据集持续增长

3.5.2 动态更新

(1) 承诺更新频率

报告对各地平台开放的数据集所承诺的更新频率进行了分析，分为不更新、不定期更新、每年更新、每半年更新、每季度更新、每月更新、每周更新、每日更新、实时更新九类。其中，报告将更新频率为不更新、不定期更新、每年或每半年的数据集视为相对静态数据，将更新频率为每季度、每月、每周、每日与实时更新的数据集视为相对动态数据。图 26 显示了各地平台数据集承诺更新频率的分布情况。按照其承诺，约 77% 的数据集为相对静态数据，这其中又以按年更新为主（占 59%）；其次为不定期更新，占 13%。承诺进行动态更新的数据集只占总数据的 23% 左右。

图 27 为各地平台数据集承诺的更新频率分布。由于部分平台未公布承诺的更新频率，因此未出现在图中。总体上，相对动态数据比例最高的城市为东莞，约半数的数据集承诺

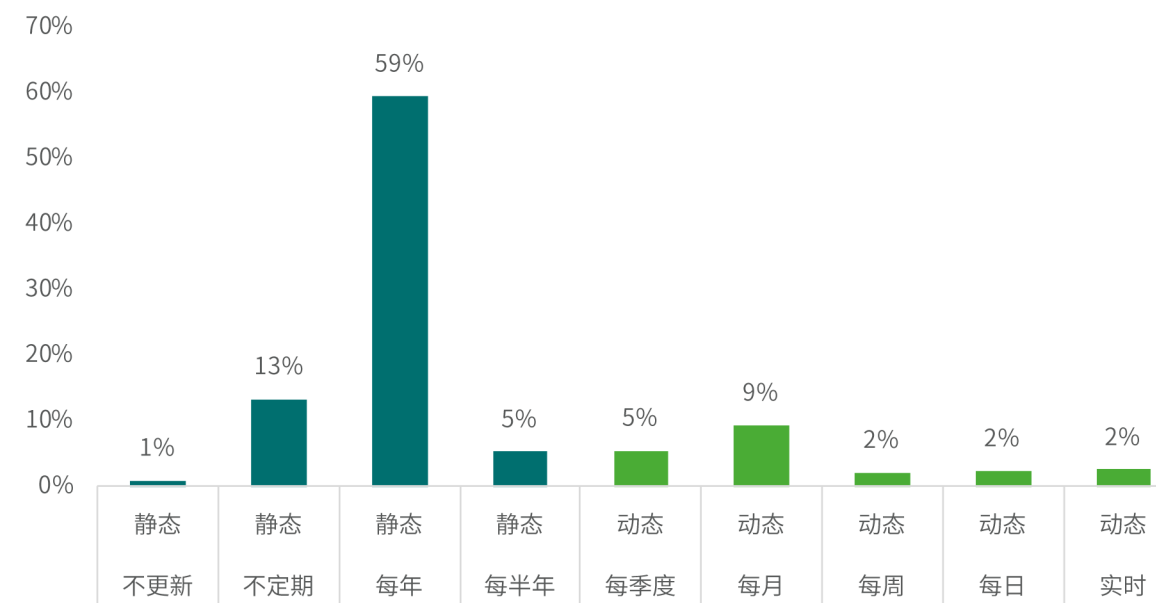


图 26 数据集承诺更新频率分布

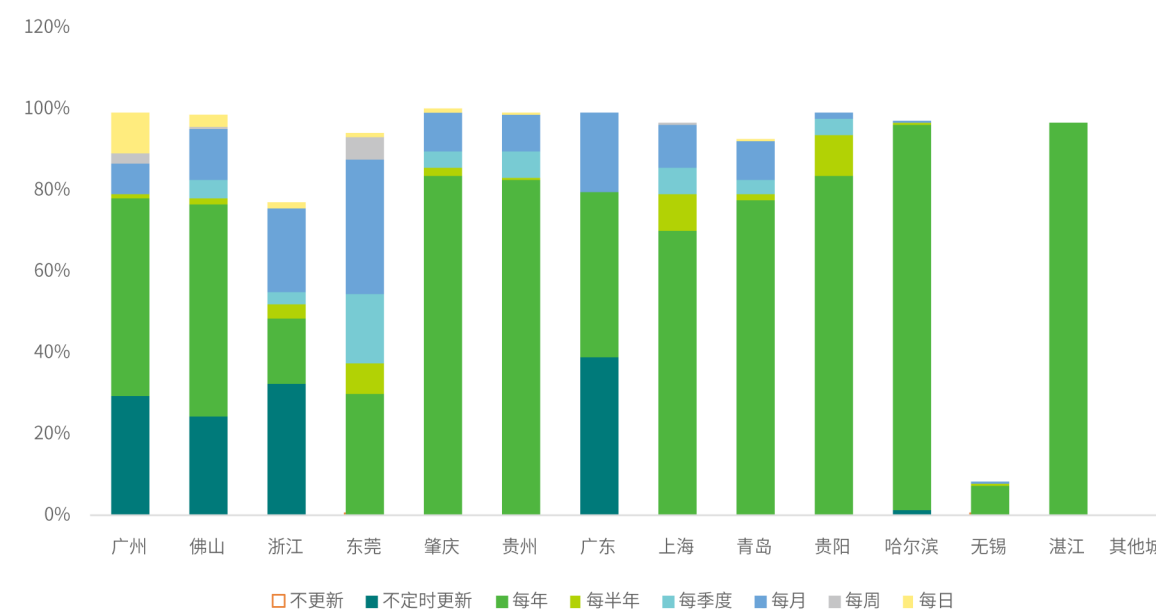


图 27 各地开放数据集承诺的更新频率分布

3. 主要发现 Key Findings

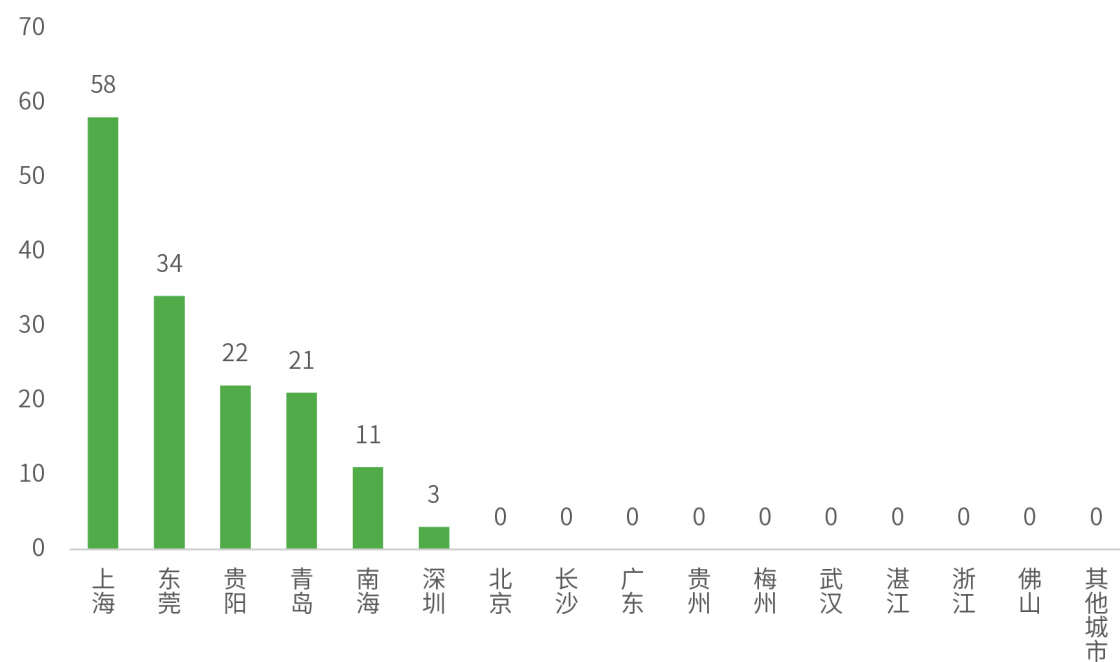


图 28 各地平台数据动态更新情况

按季度或季度以下频率更新，其中又以按月更新为主。承诺按日更新比例最高的地方分别为广州、佛山，其次是浙江、青岛、广州、上海等地，多数平台上的数据仍以承诺按年或者不定期更新为主。

(2) 实际更新情况

报告从 2017 年 4 月 18 日开始，在之后四周内的四个时间点持续采集数据，采集时间分别为 2017 年 4 月 18 日、2017 年 4 月 24 日、2017 年 4 月 30 日与 2017 年 5 月 7 日，然后通过比对数据集更新时间的变化发现在该时间内实际更新的数据集，未提供数据集更新时间的地方平台未纳入该项评估。结果如图 28 所示，上海平台上该时段内更新的数据集数量最多，其次为东莞、贵阳和青岛。

3.5.3 历史存档

历史存档是指平台将历史上不同时间更新的多个批次的数 据留存存在平台上，允许用户下载，有利于数据利用者按时间获取和利用历史数据。图 29 为实现了数据历史存档的平台分布情况，目前只有东莞、佛山、贵阳、青岛与上海五个平台。

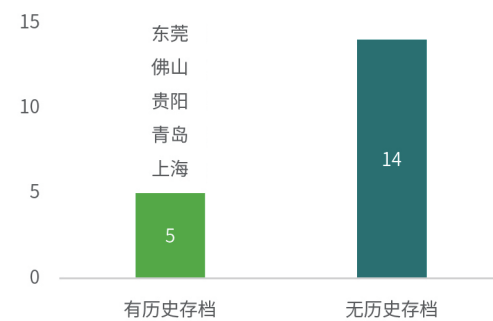


图 29 实现数据历史存档的平台分布

4. “开放数林” 指数分值与排名 Index and Ranking

四、“开放数林” 指数分值与排名

4.1 各项指数均值

各单项指标的得分均值与标准差如图 30 所示，均值得分低的指标表示各地在该项指标项上平均达标率较差。均值最低的指标项是动态更新，多数平台没有实现数据集的动态更新；其次历史存档、开放格式数量、API 数量、开放授

权；各地达标率较好的指标项主要是主题覆盖率、元数据覆盖率、关键词覆盖率等体现数据集丰富程度的指标。标准差体现的是各地平台在该项指标上得分的离散程度，标准差较高说明各地平台在该指标项上存在较大的差异。图中可见，所有单项得分的标准差都在 1 以上，这表明各地平台在所有指标项上的表现都差异较大，其中差异程度最大的是历史存档、API 数量、开放授权三项指标，既有部分平台表现优异，也有部分平台没有得分。

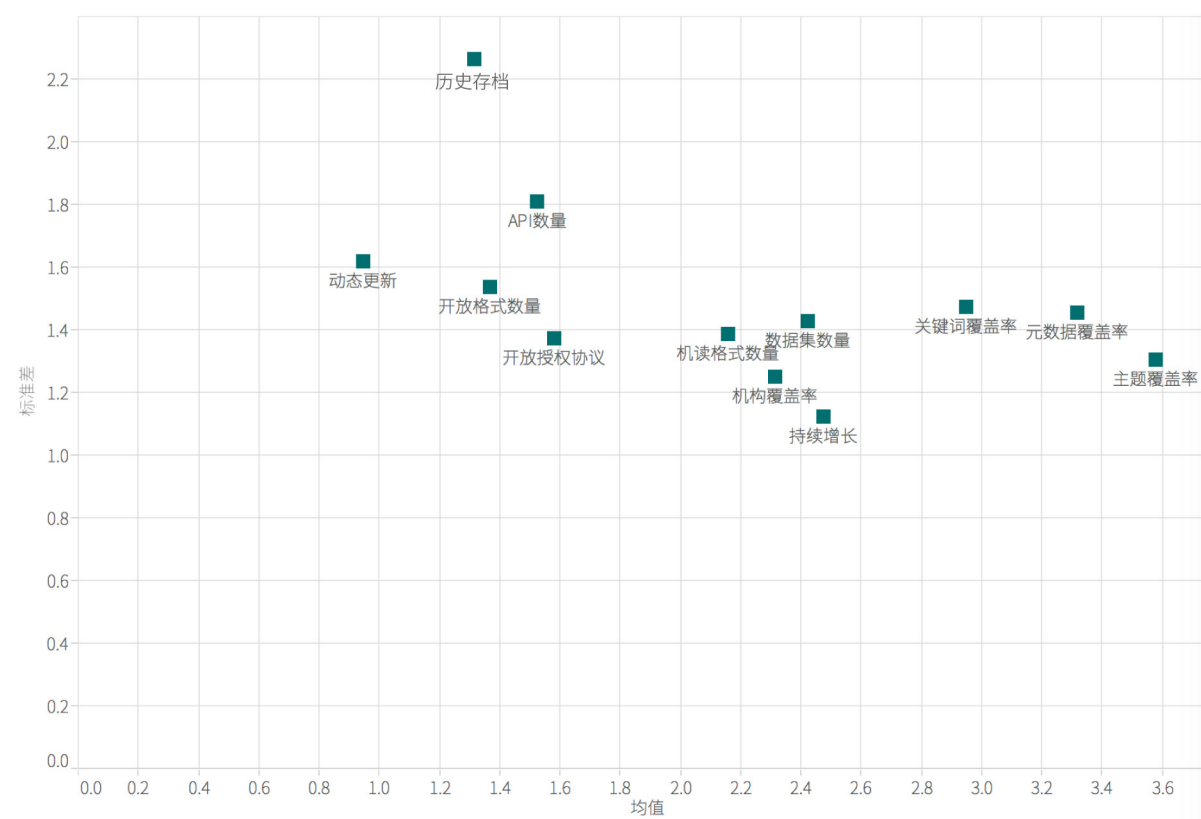


图 30 各指标得分均值与标准差分布

4. “开放数林” 指数分值与排名 Index and Ranking

4.2 指数分值与排序

各地平台的单项指标得分（0-5分）、数据层指数总分及排名如表4所示。在所有的地方平台中，表现最好的是上海、贵阳两地，其次是青岛、北京、东莞、武汉等地。在省级行政区排名中，指数得分最高的是上海，其次是北京、

广东、浙江、贵州；在副省级和地级城市排名中，得分最高的是贵阳，其次是青岛、武汉、东莞、佛山、广州、佛山南海、无锡等地，这些地方都是我国地方政府数据开放的引领者。

表4 2017 中国开放数林指数（数据层）

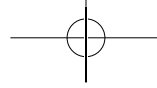
地区	数据集数量	开放授权协议	机读格式数量	开放格式数量	API数量	元数据覆盖率	机构覆盖率	关键词覆盖率	主题覆盖率	持续增长	动态更新	历史存档	指数	排名
上海	5	3.75	5	1	3	4	4	5	4	5	5	5	83.4	1
贵阳	5	3.75	5	5	2	5	5	5	5	1	3	5	83.2	2
青岛	4	2.50	4	1	2	5	3	5	5	3	3	5	69.2	3
北京	3	2.50	3	3	5	4	2	3	5	4	0	0	57.4	4
武汉	5	2.50	3	0	0	4	3	5	4	4	0	0	56.2	5
东莞	2	1.25	2	3	4	5	3	3	1	3	4	5	55.0	6
佛山	2	2.50	2	3	4	4	2	4	4	1	0	5	53.0	7
广州	2	2.50	2	3	1	5	3	3	5	2	0	0	48.6	8
南海	2	2.50	2	3	0	2	4	3	5	3	2	0	46.6	9
无锡	3	2.50	3	0	0	1	3	5	5	3	0	0	43.6	10
哈尔滨	2	0	2	2	0	5	2	3	4	2	0	0	34.2	11
深圳	3	0	2	0	5	2	2	3	3	2	1	0	33.6	12
广东	1	2.50	1	0	1	4	1	2	3	2	0	0	32.8	13
浙江	2	0	1	0	1	3	2	2	3	3	0	0	26.4	14
贵州	1	1.25	1	1	1	1	2	2	3	2	0	0	24.2	15
梅州	1	0	1	0	0	3	1	1	1	2	0	0	16.8	16
肇庆	1	0	1	0	0	1	1	1	3	3	0	0	16.4	17
长沙	1	0	1	1	0	2	0	1	3	1	0	0	15.6	18
湛江	1	0	0	0	0	3	1	1	2	1	0	0	14.6	19
其他省市	0	0	0	0	0	0	0	0	0	0	0	0	-	-

表5 2017 中国开放数林指数（省级）

地区	数据集数量	开放授权协议	机读格式数量	开放格式数量	API数量	元数据覆盖率	机构覆盖率	关键词覆盖率	主题覆盖率	持续增长	动态更新	历史存档	指数	排名
上海	5	3.75	5	1	3	4	4	5	4	5	5	5	83.4	1
北京	3	2.50	3	3	5	4	2	3	5	4	0	0	57.4	2
广东	1	2.50	1	0	1	4	1	2	3	2	0	0	32.8	3
浙江	2	0	1	0	1	3	2	2	3	3	0	0	26.4	4
贵州	1	1.25	1	1	1	1	2	2	3	2	0	0	24.2	5
其他省份	0	0	0	0	0	0	0	0	0	0	0	0	-	-

表6 2017 中国开放数林指数（副省级与地级）

地区	数据集数量	开放授权协议	机读格式数量	开放格式数量	API数量	元数据覆盖率	机构覆盖率	关键词覆盖率	主题覆盖率	持续增长	动态更新	历史存档	指数	排名
贵阳	5	3.75	5	5	2	5	5	5	5	1	3	5	83.2	1
青岛	4	2.50	4	1	2	5	3	5	5	3	3	5	69.2	2
武汉	5	2.50	3	0	0	4	3	5	4	4	0	0	56.2	3
东莞	2	1.25	2	3	4	5	3	3	1	3	4	5	55.0	4
佛山	2	2.50	2	3	4	4	2	4	4	1	0	5	53.0	5
广州	2	2.50	2	3	1	5	3	3	5	2	0	0	48.6	6
南海	2	2.50	2	3	0	2	4	3	5	3	2	0	46.6	7
无锡	3	2.50	3	0	0	1	3	5	5	3	0	0	43.6	8
哈尔滨	2	0	2	2	0	5	2	3	4	2	0	0	34.2	9
深圳	3	0	2	0	5	2	2	3	3	2	1	0	33.6	10
梅州	1	0	1	0	0	3	1	1	1	2	0	0	16.8	11
肇庆	1	0	1	0	0	1	1	1	3	3	0	0	16.4	12
长沙	1	0	1	1	0	2	0	1	3	1	0	0	15.6	13
湛江	1	0	0	0	0	3	1	1	2	1	0	0	14.6	14
其他城市	0	0	0	0	0	0	0	0	0	0	0	0	-	-



5. 结论与建议

Conclusions

五、结论与建议

1. “开放数林”来自每一棵“数木”的生长

近年来，虽然各地已陆续上线了一批政府数据开放平台，特别是《促进大数据发展行动纲要》出台后，平台增加速度明显加快。但总体上，我国绝大多数地方政府尚未开放数据。已开放政府数据的地区分布不平衡，主要分布在东南沿海地区，西部和北方地区分布较少，而且在各个行政层级上的分布都比例偏低。开放政府数据在我国刚刚起步，开放数据生态体系尚处于萌芽阶段。

建议现阶段重点以各地各级城市为切入点推动政府数据开放，鼓励有条件的城市政府基于城市需求和自身特色，率先开放政府数据，吸引社会力量对开放数据进行开发利用，创造社会价值，再反过来激励政府部门开放更多数据，构成正向闭环。同时，每一座城市的数据开放，还能形成示范效应，带动更多地方开放政府数据，共同构成未来国家政府数据开放平台的重要数据来源。此外，在鼓励各地开放数据的同时，还应促进各地开放的重点领域数据集之间的关联融合。独木不成林，一人不为众，只有每一棵“数木”既竞相相生，各有千秋，又枝杈相连，能量互通，才能长出一整座开放数据的森林。

2. 开放真正能“用”的数据

目前各地开放数据平台上线的数据集在数量上已成规模，多个地方的数据集总数已突破或接近 1000，但其中能真正符合完整的、原始的、可机读的、非专属的、以接口形式提供等开放数据标准的数据集仍然偏少。不符合标准要求的所谓“数据集”却比比皆是，或为加工归总后的统计报表，或为非结构化、不可机读的文本内容，或为拆分后或未整合的单行数据，甚至还有有的数据集名称下面不存在可获取的数据集。这些“数据集”无法被利用，更不可能产生价值，将

会使数据开放最终流于形式。

在开放数据集的同时提供全面的元数据信息有助于用户理解和应用数据。目前，大多数地方平台都能提供基本的元数据，但各地情况参差不齐，更新频率、访问量、数据项和数据量等条目的覆盖率仍有待完善。

建议对主管部门加强有关开放数据基本知识的培训，厘清概念，达成共识；然后基于开放数据的基本原则和标准，从数据利用者的视角和需求出发，开放完整的、原始的、可机读的、非专属的、结构化的、高价值的数据集，并配备实用详实的元数据条目，帮助用户用好数据、用对数据。物尽其用，数尽其利，只有开放真正能被“用”起来的数据才能创造价值。

3. “授予”用户利用数据的权利

目前，许多地方的政府数据开放平台还未提供明确充分的数据开放授权，来确保用户免费获取、不受歧视、自由利用、自由传播和分享数据的权利，即所谓开放数据的法律性开放。建议尽快制定相关法律法规，明确界定“开放数据”和“不开放数据”的边界，制定相应的正面清单和负面清单，在保障国家秘密、商业秘密和个人隐私的前提下，最大限度地开放政府数据，对于法律允许开放的数据应明确保障任何人不受限制地对其进行自由利用和自由分享的权利。有舍才有得，只有被充分“授权”的数据，才能被安心利用，并释放能量。

4. 让数据“活”起来

自平台上线后，只有少数地方能保持数据集持续增长，不少地方出现中断甚至停滞。即使是已开放的数据集，真正实现动态更新的比例也极低。此外，只有少数地方将历史上多个批次的数据留存在平台上，允许用户下载。“流水不腐，

户枢不蠹”，建议建立常态化的工作机制，确保开放数据集存量动态更新，增量持续不断，历史量多批次保存。“问渠那得清如许？为有源头活水来”，只有源源不断的数据供给，才有可能真正激发创新活力。

5. 让“用数据的人”有获得感

各地平台数据集在主题领域设置、数据提供部门选择和数据集名称设置等方面，都更多体现了政府部门的视角，而非用户视角。社会公众是否看得懂政府部门设置的数据主题分类和数据集？能否能找到他们所需要的数据？社会最需要的数据是否就在提供数据最多的部门那里？政府开放最多的数据集就一定是用户最想要的吗？以人为本，而不是“本人以为”。建议在制定政府数据开放规划和决策时多从用户的实际需求而非政府部门的自我判断出发，不断向数据利用者征集需求和建议，才能更有针对性地供给真正高需求、高价值的数据集，让“用数据的人”有更多获得感。



附录：平台数据集主题分布

Appendix

经济工商	交通出行	机构团体	文化休闲	医疗卫生	教育科技	社会民生	资源环境	城建住房	公共安全	农业农村	社保就业	财政金融	信用服务	其他
经济工商 经济建设58 投资服务1 企业服务50	交通出行 交通服务35	机构团体 政府与社会团体46 企业2 民办非企业6	文化休闲 餐饮美食7 旅游住宿20 文体娱乐32 消费购物5	医疗卫生 医疗保障45	教育科技 教育科研16 教育科研16	社会民生 生活服务30	资源环境 环境与资源保护29	城建住房 房屋住宅5	公共安全 公共安全6	农业农村 农村居民3 农业农村70	社保就业 劳动就业13 社会保障5	财政金融 金融信息13 社会保障5	信用服务 宗教信仰7	其他 资质认定267 红榜34 黑榜37
佛山 经济建设37	道路交通11	机构团体40	文化休闲15	卫生健康20	教育科技43	民生服务46	社会资源79	城市建设24	公共安全23					
广东 经济建设15	道路交通18	机构团体18	文化休闲13	卫生健康3	教育科技7	社会发展12	资源环境16		公共安全11					
广州 经济建设36	道路交通9	机构团体18	文体娱乐24	卫生健康34	教育科技20	民生服务17	资源环境21	城市建设17	公共安全1	农业农村8	劳动人事5	财政金融8		
贵阳 经济建设128 企业服务77	交通运输60	政府机构62	文化休闲110	卫生健康81	教育科技67	生活服务32	生态环境83		公共安全135	三农服务58	劳动人事41	财政金融85	信用服务16	
贵州 经济建设31 建材信息1	交通运输1	政府机构53 政务信息1	文化休闲1	医疗健康12	科技教育12	社会发展18	环境资源11	城市建设14	防灾减灾1				信用服务1	文普信息1 新闻信息2 综合信息1
哈尔滨 商事信息39	道路交通6	机构团体57	文体娱乐8	卫生健康10	教育科技4	民生服务63	资源环境11	城市建设13						统计数据6
梅州														
南海 企业服务77	交通服务34	政府机构与社会团体24	餐饮服务6 旅游住宿8 文体娱乐18 消费购物5	医疗健康29	教育科研21	生活服务32	环境与资源保护19	城市建设22	生活安全3	农业农村12	社会保障16			宗教信仰6
青岛 经济人才7 经济主体59 经济指数47	交通监管2 交通资讯44 交通运输5	公共机构9 公用事业16 社会团体4 法律组织9	文化休闲8 休闲娱乐11 体育建设4	卫生监督26 医疗卫生58 动物诊疗7	教育培训37 科技教育14	便民服务29 民生服务35	节能环保16 自然资源14	城市管理4 城市建设37 住房保障6	安全生产8 防灾减灾4 公共安全13	农业畜牧19 水利建设9	劳动就业15 社会保障7	金融服务9 财政金融9 国地税局5 预算执行1		工作结果1 结果公告1 项目计划1 行政处罚1 行政执行2 市场监管71 宗教信仰15 法律执业3
上海 经济建设283	道路交通131	机构团体51	文化休闲64	卫生健康110	教育科技93	民生服务98	资源环境78	城市建设126	公共安全35				信用服务13	数据领域6
深圳 经济建设123	道路交通14		文化娱乐52	公共卫生9 医疗卫生34 人口计划12	公共教育14	社会发展52		城市建设128	公共安全4		就业服务20 社会保障17			
无锡 企业服务272	交通服务11	政府机构与社会团体53	100	医疗健康20	教育科技30	生活服务51	环境与资源保护4	房屋住宅5	生活安全1	农业农村9	劳动就业2 社会保障25			信息公开8 宗教信仰4
武汉 经济基础382	交通服务69	政府机构196	0	医疗卫生42	教育科技82	公共服务277	能源环境104	公共安全17	公共安全17	农业农村11		金融服务35	法律服务23	
湛江 工商5	交通运输3	公安12						房产4	公安12					统计1
长沙 工业1	交通6		文化工业5	医疗卫生1	服务8 民生13	能源3	能源3	住宅1		林业2 农业3		金融4		
肇庆 经济建设12	道路交通4		文化休闲5	卫生健康8	教育科技9	社会发展8	资源环境14	城市建设8	公共安全5					
浙江 经济建设273	道路交通7	机构团体15	文化休闲29		教育科技14	民生服务7	环境资源6	城市建设10						

参考文献

Bibliography

- 郑磊 (2015) . 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 (2015) . 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.
- 郑磊、熊久阳 (2017) . 中国地方政府开放数据研究：技术与法律特性. 公共行政评论，2017-1: 53-74

关于“开放数林”

About China Open Data Index

开放数林

中国开放数林指数

CHINA OPEN DATA INDEX

· 2017 ·

2012年以来，我国已有近二十个地方政府陆续推出数据开放平台。开放数据，蔚然成林。一棵棵开放“数木”的丛然并生、成荫如盖、直至叶茂花开，终将成长为一片繁盛多样、平衡稳定、循环持续的“开放数林”。“中国开放数林指数”将持续追踪我国地方政府数据开放工作的进展和问题，精心测量“数木”们的形态、密度、材质与价值，助力中国开放数据生态系统的形成与发展。

联合发布方



复旦大学



提升政府治理能力大数据应用技术国家工程实验室

指数出品方



DMG Lab
Fudan University
复旦大学 | 数字与移动治理实验室



扫描二维码关注
复旦大学数字与移动治理实验室
微信公众号

指数出品团队

郑磊、吕文增、高丰、熊久阳、赵丽琼、刘新萍、关文雯、蔡城城、任雅丽、纪昌秀

合作伙伴



DESIGN
麦明德设计事务所

冥睿(上海)
信息科技有限公司

专家评委会名单

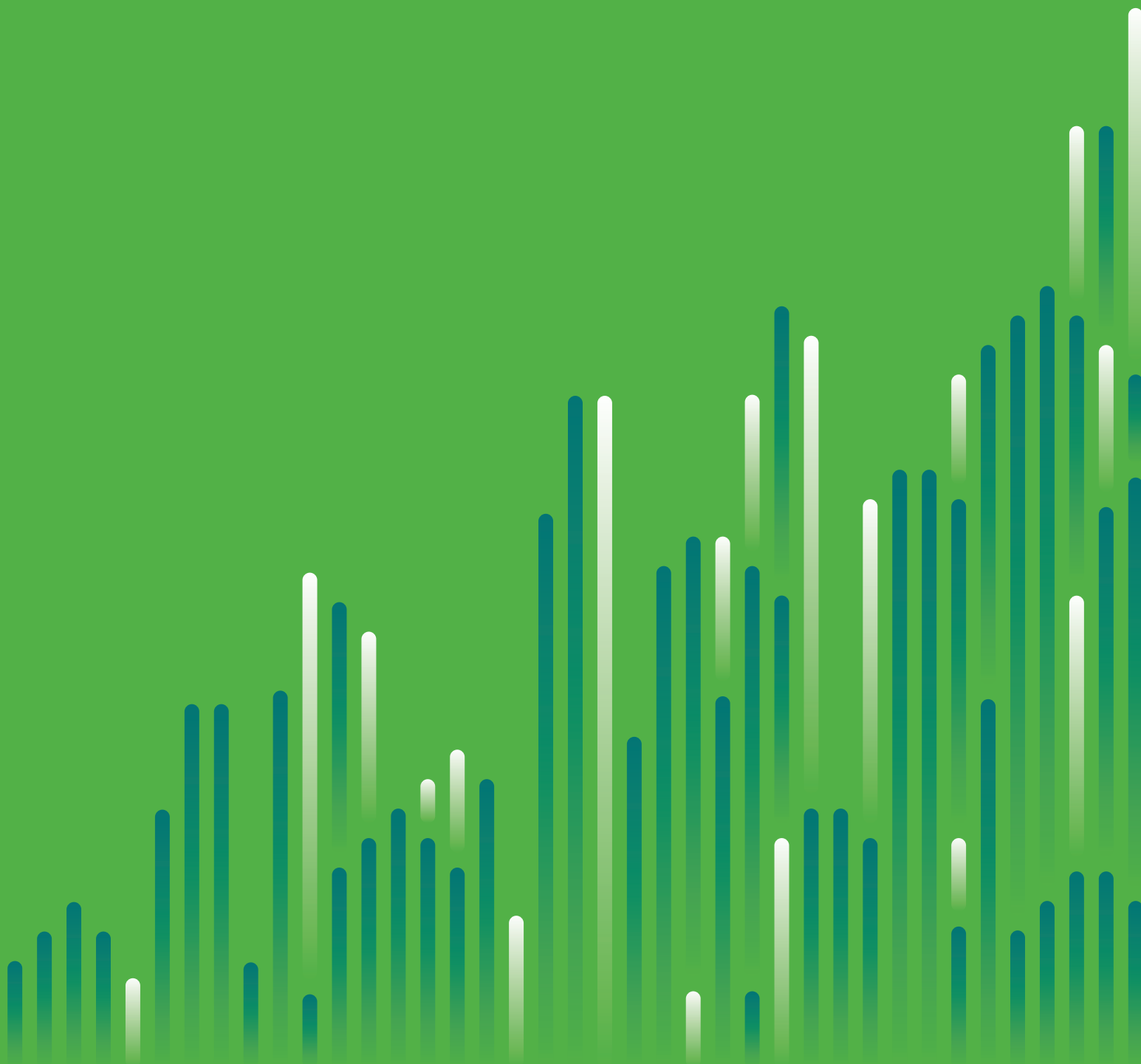
Advisory Committee

姓名	单位	姓名	单位
陈涛	华中科技大学	马小峰	同济大学
陈瑶	上海对外经贸大学	孟天广	清华大学
陈玉梅	暨南大学	宋文好	《电子政务》
邓崧	云南大学	孙未未	复旦大学
樊博	上海交通大学	汤志伟	电子科技大学
范静	北京外国语大学	涂子沛	杭州涂子沛数据科技咨询有限公司
高丰	开放数据中国	万如意	北京众标智能科技有限公司
韩涵	中国信息通信研究院	王法硕	华东师范大学
胡辟砾	独立咨询师	王芳	南开大学
黄璜	北京大学	王建冬	国家信息中心
黄如花	武汉大学	王鹏	清华同衡
黄志敏	数据工场	王少辉	武汉大学
金耀辉	上海交通大学	王咏笑	城市数据团
李伯一	英国埃克塞特大学	夏义堃	武汉大学
李谦升	上海大学	肖卫兵	上海政法学院
李雯	城市数据派	熊竞	上海市民政科研中心
李重照	云南财经大学	薛梓闻	上海科巴度数据管理有限公司
利嘉豪	广州东西网络科技有限公司	叶超	百度（中国）有限公司
林佑达	城市策略顾问	张峰	浪潮集团
刘春蕾	上海闵行区青悦环保信息技术服务中心	张楠	清华大学
刘文静	暨南大学	章燕华	浙江大学
刘啸	上海数慧系统技术有限公司	郑跃平	中山大学
刘新萍	上海理工大学	朱京	上海美华系统有限公司
龙怡	上海政法学院	朱琳	华东理工大学
马亮	中国人民大学		

注：按姓氏拼音排序



DMG Lab
Fudan University
复旦大学 | 数字与移动治理实验室



开放数据 蔚然成林

