



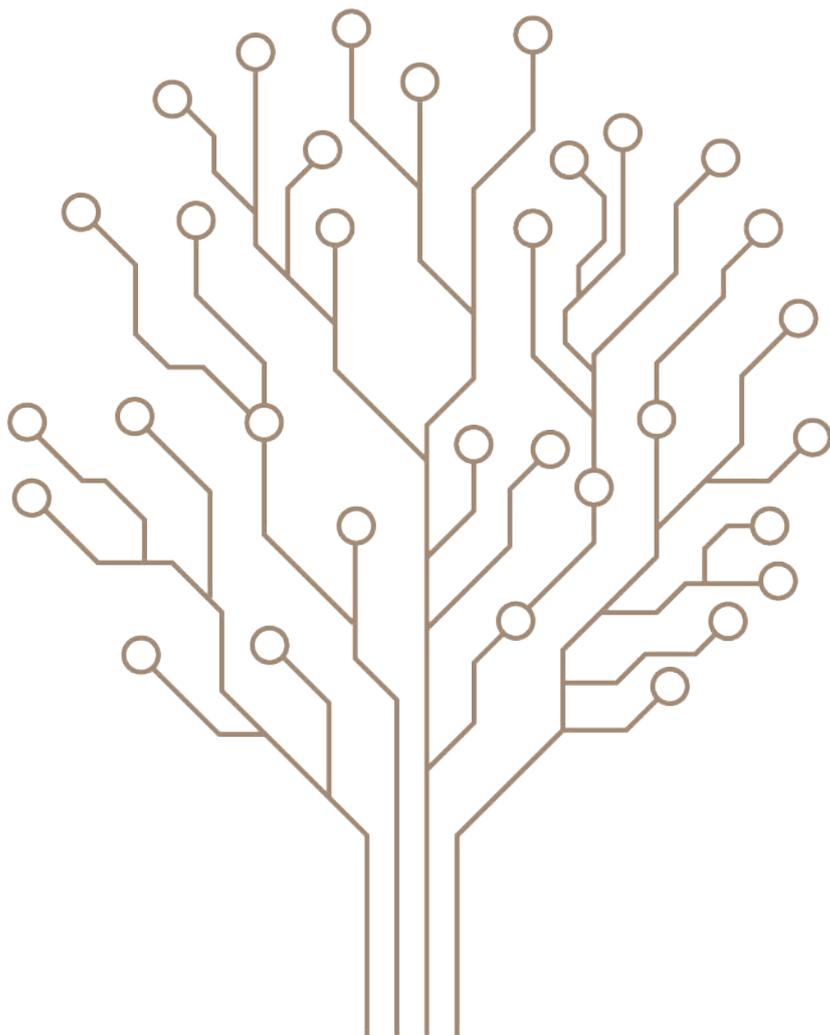
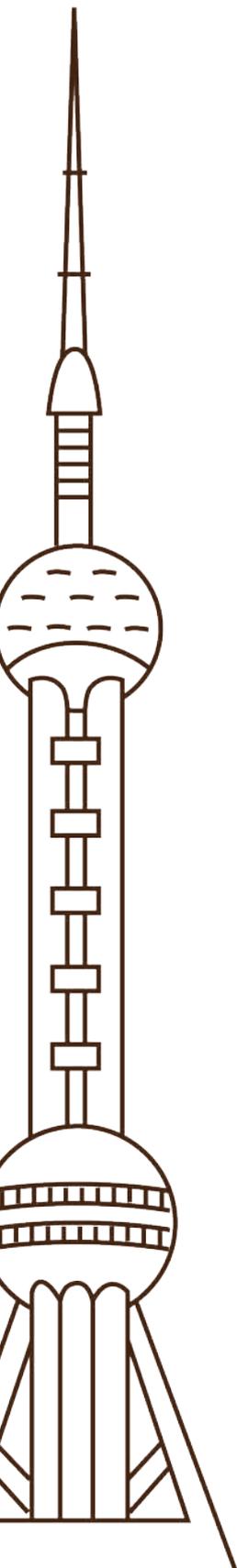
中国开放数林指数
CHINA OPEN DATA INDEX

2019

上半年

上海开放数林指数分析报告

复旦大学数字与移动治理实验室



CONTENTS

■ 评估方法	3
1.1 评估指标体系	3
1.2 数据采集分析方法	4
1.3 得分计算方法	5
■ 分值与建议	6
1 准备度	6
1.1 法规与政策	6
1.2 组织与执行	7
1.3 标准规范	8
1.4 建议	8
2 平台层	9
2.1 数据发现	9
2.2 数据获取	10
2.3 工具提供	11
2.4 利用展示	11
2.5 互动反馈	12
2.6 公众传播	13
2.7 账户体验	13
2.8 建议	13
3 数据层	14
3.1 数据数量	15
3.2 数据质量	15
3.3 数据标准	17
3.4 数据覆盖面	18
3.5 数据持续性	19
3.6 建议	19
4 利用层	20
4.1 利用促进	20
4.2 成果产出	20
4.3 数据利用	22
4.4 建议	22

评估方法

1.1 评估指标体系

■ 评估指标体系共包括准备度、平台层、数据层、利用层四个维度及下属多级指标（见图1）：

■ 准备度是“数根”，是数据开放的基础，包括法规与政策、组织与执行、标准规范等三个一级指标。

■ 平台层是“数干”，是政府开放数据和用户获取数据的载体，是展示利用成果的中心，也是连接数据开放供给侧和需求端的桥梁，包括数据发现、数据获取、工具提供、利用展示、互动反馈、公众传播和账户体验等七个一级指标。

■ 数据层是“数叶”，是数据开放的核心，包括数据数量、数据质量、数据标准、数据覆盖面、数据持续性等五个一级指标。

■ 利用层是“数果”，是数据开放的成果。包括利用促进、成果产出和数据利用等三个一级指标。

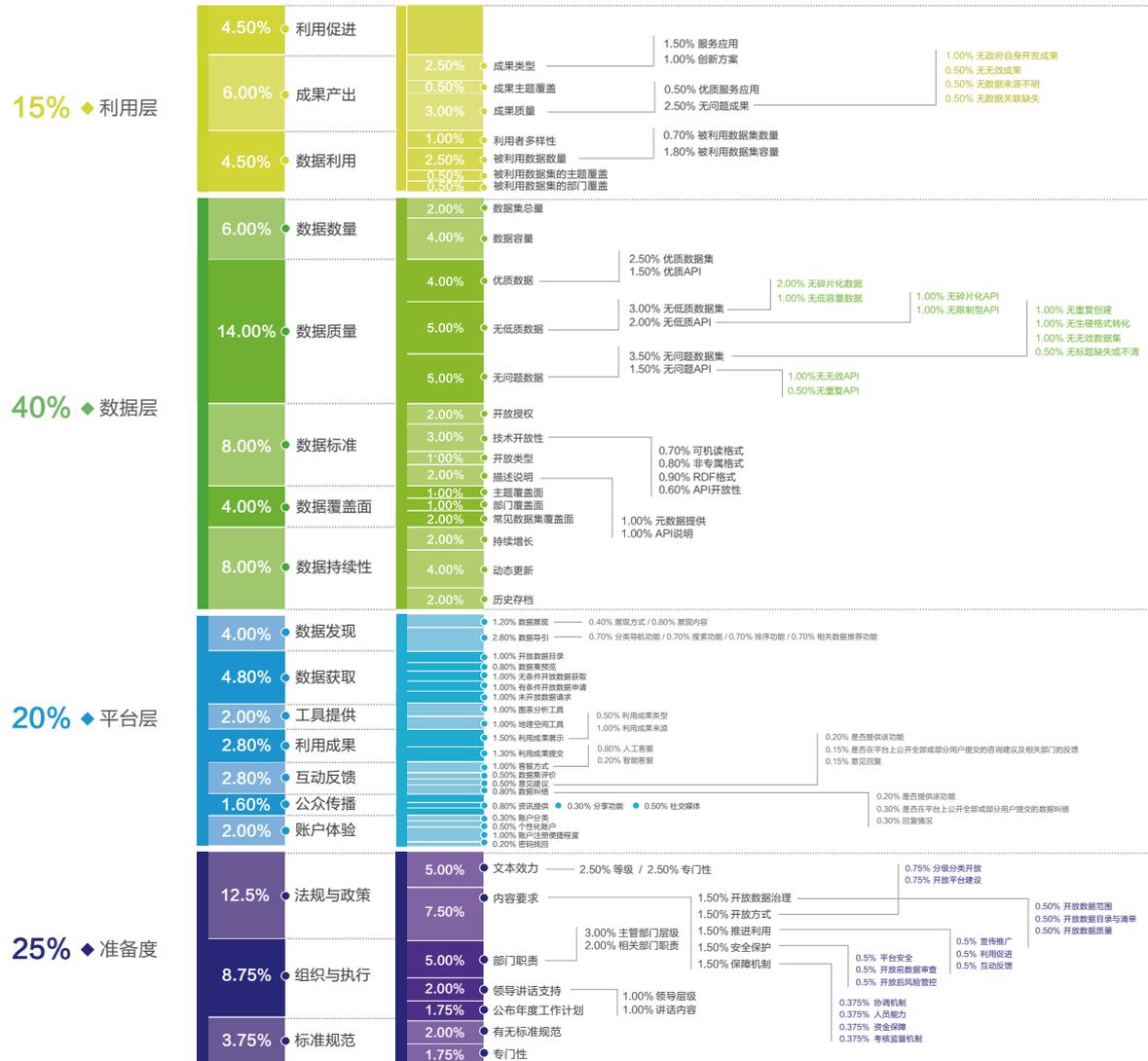


图 1 评估指标体系

1.2 数据采集分析方法

■准备度评估主要对相关法律法规、政策、意见、计划、标准规范、新闻报道等资料进行了描述性统计分析和文本分析。搜索方法主要包括以下两种：一是在百度搜索引擎以关键词检索相关法规与政策文本、年度工作计划、有关地方党政领导讲话支持的新闻报道以及数据开放主管部门的信息；二是在上海政府门户网站以及政府数据开放平台上通过人工观察和关键词检索采集数据。指数出品方对法律法规与政策的筛选标准为至少有一节专门针对政府数据开放的内容；对标准规范的筛选标准为有关政府数据开放的地方性标准，行业性的标准规范未纳入本次评估范围。数据采集截止时间为 2019 年 4 月 15 日。

■平台层评估主要采用人工观察法对各地方政府数据平台上各项功能进行观测并做描述性统计分析，数据采集截止时间为 2019 年 3 月 31 日。同时，指数出品方还对平台的回复情况（包括回复时间和回复质量）进行了评估，回复情况采集截止时间为 2019 年 4 月 4 日。

■数据层评估主要通过机器自动抓取各地平台上开放的数据结合人工观察采集相关信息，然后对数据进行了描述性统计分析、交叉分析、文本分析和空间分析。数据采集截止时间为 2019 年 4 月 15 日，对“动态更新”这一指标的评测时段为 2019 年 1 月 1 日至 2019 年 4 月 15 日。

■利用层评估主要对上海平台上展示的利用成果进行了人工观察和测试，对 2017 年以来各地开展的开放数据创新利用比赛信息进行了网络检索，并对采集到的数据进行了描述性统计分析。数据采集截止时间为 2019 年 4 月 4 日。

1.3 得分计算方法

■基于上海在各项评估指标上的实际表现从低到高按照 0-5 分共 6 档分值进行评分，其中 5 分为最高分，相应数据缺失或完全不符合标准则分值为 0。对于连续型统计数值类数据则使用极差归一法将各地统计数据结果换算为 0-5 之间的数值作为该项得分。

■报告对应各项评估指标，以处于全国前 10%、前 25%、前 40%、前 70% 和后 30% 五个等级来展示上海在该项指标上所处的位置。

■ 分值与建议

1. 准备度

上海在准备度各项指标上的得分如图 2 所示。

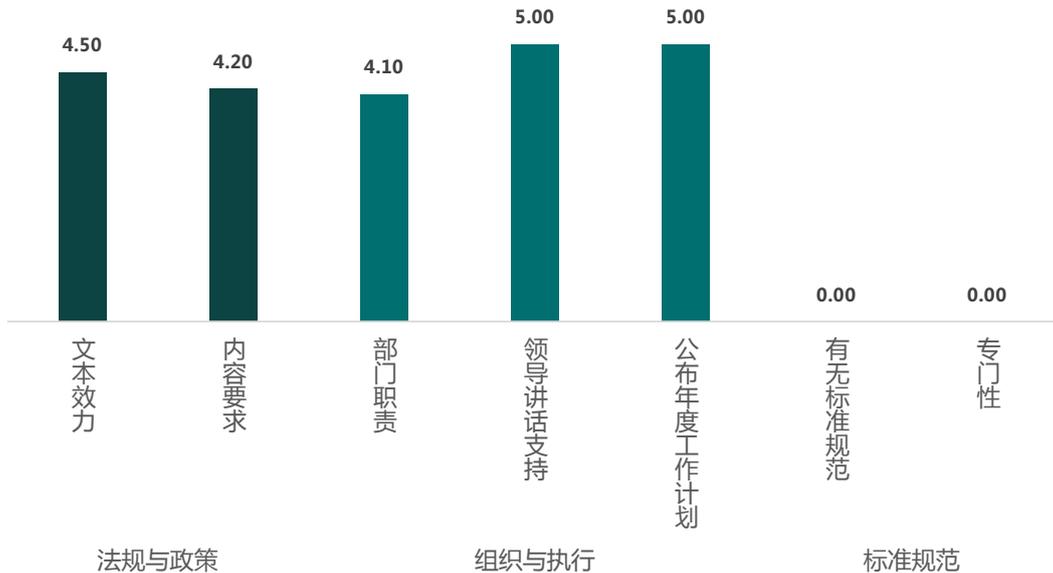


图 2 上海准备度各级指标得分

1.1 法规与政策

指对政府数据开放相关环节（如开放数据治理、开放方式、推进利用、安全保护以及保障机制）作出规范性要求的法律法规和政策文件，包括文本效力评估和内容要求分析两部分。

(1) 文本效力

指法规和政策文件所具有的约束力，主要表现在等级和专门性两方面。

- 等级：上海在此项排名居于所有在评地区的前 10%；
- 专门性：上海在此项排名居于所有在评地区的前 10%。

(2) 内容要求

对纳入评估范围的法规与政策中的内容进行了分析，内容包括开放数据治理、开放方式、推进利用、安全保护以及保障机制五个方面的要求。

● 开放数据治理

开放数据范围：上海在此项排名居于所有在评地区的前 10%；

开放数据目录与清单：上海在此项排名居于所有在评地区的前 10%；

开放数据质量：上海在此项排名居于所有在评地区的前 20%。

● 开放方式

分级分类开放：上海在此项排名居于所有在评地区的前 10%；

开放平台建设：上海在此项排名居于所有在评地区的前 10%。

● 推进利用

宣传推广：上海在此项排名居于所有在评地区的前 10%；

利用促进：上海在此项排名居于所有在评地区的前 10%；

互动反馈：上海在此项排名居于所有在评地区的前 10%。

● 安全保护

平台安全：上海在此项排名居于所有在评地区的后 30%；

开放前数据审查：上海在此项排名居于所有在评地区的前 10%；

开放后风险管控：上海在此项排名居于所有在评地区的后 30%。

● 保障机制

协调机制：上海在此项排名居于所有在评地区的前 10%；

人员能力：上海在此项排名居于所有在评地区的前 10%；

资金保障：上海在此项排名居于所有在评地区的前 10%；

考核监督机制：上海在此项排名居于所有在评地区的前 10%。

1.2 组织与执行

指开放数据的组织保障与执行推进，包括部门职责、领导讲话支持以及公布年度工作计划三部分。

(1) 部门职责：包括政府数据开放主管部门及相关部门的职责分工。

● 主管部门层级：上海在此项排名居于所有在评地区的后 30%；

● 相关部门职责：上海在此项排名居于所有在评地区的前 10%。

(2) 领导讲话支持：指地方党政高层领导公开讲话对政府数据开放工作做出明确要求和支持。

● 领导层级：上海在此项排名居于所有在评地区的前 10%；

● 讲话内容：上海在此项排名居于所有在评地区的前 10%。

(3) 公布年度工作计划：上海在此项排名居于所有在评地区的前 10%。

1.3 标准规范

指制定数据开放的评价标准和操作规范。

● 有无标准规范：上海在此项排名居于所有在评地区的后 30%；

● 专门性：上海在此项排名居于所有在评地区的后 30%。

1.4 建议

根据准备度层各项指标上，上海在全国所处的位置，提出以下建议：

(1) 完善数据开放工作的安全保障机制。

(2) 制定专门针对数据开放的标准规范。

(3) 加强体制机制保障，增强主管部门的工作协调能力。

(4) 在文本效力等级方面进一步提高，通过地方性法规为数据开放提供强有力的法治保障。

2. 平台层

上海在平台层各项指标上的得分如图 3 所示。

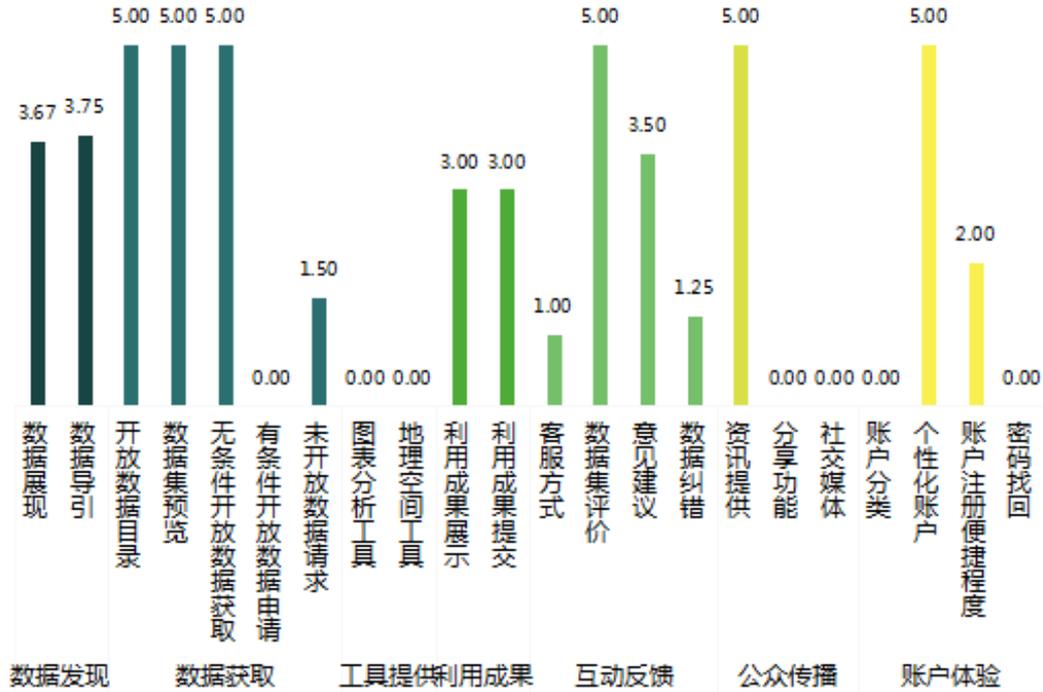


图 3 上海平台层各级指标得分

2.1 数据发现

政府数据开放平台以醒目的方式提供数据展现和数据导引。

(1) 数据展现：平台呈现的开放数据概况、动态以及展现数据的方式。

● 展现内容：平台在首页醒目位置展现了开放数据概况和数据动态。

数据概况：平台上展示的数据的基本概况。

上海在此项排名居于所有在评地区的前 70%。

数据动态：平台上展示的最新数据情况。

上海在此项排名居于所有在评地区的前 10%。

● 展现方式：对平台上开放数据的概况和动态以各种直观、形象的方式进行展现。

上海在此项排名居于所有在评地区的前 40%。

(2) 数据导引：平台提供了便捷的导引功能，以帮助用户发现所需要的目标数据。

- 分类导航功能：平台提供了按主题、部门、标签和场景等类别进行数据导引的功能。

上海在此项排名居于所有在评地区的前 10%。

- 搜索功能：平台提供了按名称搜索、字段搜索或其他条件搜索数据集的功能。

上海在此项排名居于所有在评地区的前 10%。

- 排序功能：平台提供了按照某项具体标准对数据集进行的排序。

上海在此项排名居于所有在评地区的前 10%。

- 相关数据推荐功能：在某个数据集下推荐与之相关联的数据集。

上海在此项排名居于所有在评地区的后 30%。

2.2 数据获取

政府数据开放平台提供便于用户浏览和获取目标数据的相关功能。

(1) 开放数据目录：平台提供的所有已开放数据资源的完整目录。

上海在此项排名居于所有在评地区的前 10%。

(2) 数据集预览：用户在获取数据集之前，能够对数据集进行预览以了解其详细内容。

上海在此项排名居于所有在评地区的前 10%。

(3) 无条件开放数据获取：用户获取无条件开放类数据集的方式。

上海在此项排名居于所有在评地区的前 10%。

(4) 有条件开放数据申请：用户获取有条件开放类数据集的方式。

上海在此项排名居于所有在评地区的后 30%。

(5) 未开放数据请求：用户对于尚未开放的数据集，可通过平台提出数据开放请求。

- 提供该功能：上海在此项排名居于所有在评地区的前 10%。

- 展示用户数据请求：上海在此项排名居于所有在评地区的后 30%。

- 平台回复情况（时效和质量）：上海在此项排名居于所有在评地区的后 30%。

2.3 工具提供

政府数据开放平台提供了能够帮助初级用户对数据集进行分析和利用的基本工具。

(1) 图表分析工具：平台提供了数据可视化或统计分析的工具。

上海在此项排名居于所有在评地区的后 30%。

(2) 地理空间工具：平台提供了将位置类数据叠加在地图上进行展现的功能。

上海在此项排名居于所有在评地区的后 30%。

2.4 利用展示

在政府数据开放平台上集中展现用户利用平台上开放的数据产生的各类数据利用成果。

(1) 利用成果展示：对数据利用成果类型和利用成果来源的展示。

●利用成果类型：对基于开放数据产生的利用成果的类型展示。

上海在此项排名居于所有在评地区的前 10%。

●利用成果来源：对基于开放数据产生的利用成果的来源展示。

上海在此项排名居于所有在评地区的前 10%。

(2) 利用成果提交：平台为利用者开设了提交数据利用成果的功能。

上海在此项排名居于所有在评地区的前 25%。

2.5 互动反馈

政府数据开放平台提供了便于用户与平台或数据提供方进行沟通反馈的功能。

(1) 客服方式：目前主要有人工客服和智能客服两种。

●人工客服：用户能够通过平台上提供的联系方式与平台或数据发布方联系。

平台联系方式：上海在此项排名居于所有在评地区的前 10%。

数据发布者联系方式：上海在此项排名居于所有在评地区的后 30%。

●智能客服：平台提供了人机智能交互的方式来为用户提供个性化的指引和服务。

上海在此项排名居于所有在评地区的后 30%。

(2) 数据集评价：用户可对单个数据集进行的星级评分和文字评价的功能。

●提供该功能：上海在此项排名居于所有在评地区的前 10%。

●展示用户评价：上海在此项排名居于所有在评地区的前 10%。

(3) 意见建议：平台通过问卷调查和提交文字内容等方式向用户采集意见或建议。

●提供该功能：上海在此项排名居于所有在评地区的前 10%。

●展示用户意见建议：上海在此项排名居于所有在评地区的前 10%。

●平台回复情况（时效和质量）：上海在此项排名居于所有在评地区的后 30%。

(4) 数据纠错：用户可向平台反馈其发现平台中数据集存在的问题或错误。

●提供该功能：上海在此项排名居于所有在评地区的前 10%。

●展示用户数据纠错：上海在此项排名居于所有在评地区的后 30%。

●平台回复情况（时效和质量）：上海在此项排名居于所有在评地区的后 30%。

2.6 公众传播

政府通过开放平台或社交媒体等渠道向社会进行传播。

(1) 资讯提供：通过开放平台提供与数据开放相关的资讯。

上海在此项排名居于所有在评地区的前 10%。

(2) 分享功能：在平台上提供信息分享传播功能，使用户可将平台上的信息在微博、微信和 QQ 等社交媒体平台进行转发传播。

上海在此项排名居于所有在评地区的前 10%。

(3) 社交媒体：开设专门的社交媒体账号，有助于向社会公众宣传推介本地的数据开放工作并与社会公众通过社交媒体开展互动。

上海在此项排名居于所有在评地区的后 30%。

2.7 账户体验

政府数据开放平台为用户在开设和使用账户时提供便捷的个性化服务。

(1) 账户分类：平台不仅为自然人，还为法人和非法人组织提供了注册功能。

上海在此项排名居于所有在评地区的后 30%。

(2) 个性化账户：平台为用户提供个性化的账户功能。

上海在此项排名居于所有在评地区的前 10%。

(3) 账户注册便捷程度：用户在平台注册账户的便捷程度。

上海在此项排名居于所有在评地区的前 70%。

(4) 密码找回：用户可以通过提交申请，完成相关验证后即可修改或找回账户密码。

上海在此项排名居于所有在评地区的后 30%。

2.8 建议

根据平台层各项指标上，上海在全国所处的位置，提出以下建议：

(1) 在已展示的数据集下推荐相关的数据集；

(2) 回复用户对有条件开放数据的申请；

(3) 及时回复并公开展示用户对未开放数据的请求、对数据的纠错反馈以及对平台的意见建议；

(4) 提供帮助用户对数据集进行分析和开发的基本工具；

(5) 展示数据发布者的联系方式；

(6) 提供企业或单位法人代表的注册方式，增设密码找回功能，考虑智能服务和社交媒体的开设。

(7) 进一步增加对平台数据概况内容的展示，降低用户注册的难度，丰富数据的展现方式，增加提交其他利用成果的渠道。

3. 数据层

上海在数据层各项指标上的得分如图 4 所示。

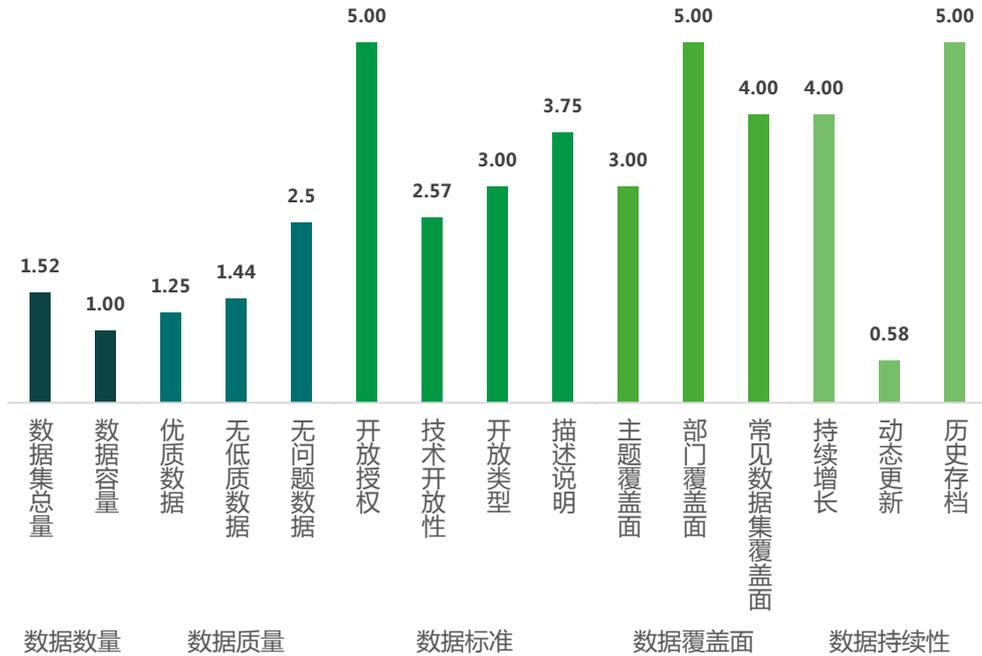


图 4 上海数据层各级指标得分

3.1 数据数量

(1) 数据集总量：数据集是由数据组成的集合，通常以表格形式出现，每一“列”代表一个特定变量，每一“行”则对应一个样本单位。

上海在此项排名居于所有在评地区的前 25%。

(2) 数据容量：数据容量是将一个地方平台上可下载的、结构化的、各个时间批次的数据集的字段数（列数）乘以条数（行数）后得出的数据总量。

上海在此项排名居于所有在评地区的前 40%。

3.2 数据质量

(1) 优质数据

● 优质数据集：数据容量大，社会需求高的是优质数据集。

上海在此项排名居于所有在评地区的前 25%。

● 优质 API：指的是地方平台上提供的必要、可调取真实数据、至少每天动态更新、数据集容量较大的 API 接口。

上海在此项排名居于所有在评地区的后 30%。

(2) 无低质数据

● 无低质数据集

无碎片化数据：碎片化数据是指按照时间、行政区划、政府部门、批次等被人为分割的数据集。

上海在此项排名居于所有在评地区的前 70%。

无低容量数据：低容量数据是指条数在三行或三行以内的数据集。

上海在此项排名居于所有在评地区的前 70%。

● 无低质 API

无碎片化 API：碎片化 API 是指通过 API 接口提供的数据集按照时间、行政区划、政府部门、批次等被人为分割。

上海在此项排名居于所有在评地区的前 70%。

无限制型 API：限制型 API 是指平台对于数据容量较小、更新频率低的数据集仅提供了通过 API 接口获取一种方式，而未提供下载获取的方式。

上海在此项排名居于所有在评地区的后 30%。

(3) 无问题数据

● 无问题数据集

无重复创建：重复创建是指平台上出现标题相同、可下载数据文件内容相同、所属主题也相同的数据集。

上海在此项排名居于所有在评地区的后 30%。

无生硬格式转化：生硬格式转化是指平台将非结构化的 DOC、PDF 等文件生硬地转化成 XLS、CSV、XML 等可机读格式。

上海在此项排名居于所有在评地区的前 10%。

无无效数据集：无效数据集包括以下三类情况：数据集名称下没有可获取的数据集；只提供数据链接，却无法获取数据集；数据集内实际上不存在数据。

上海在此项排名居于所有在评地区的前 10%。

无标题缺失或不清晰：标题缺失或不清晰是指平台上开放的数据集标题不全，或是标题名称表达不清、不知所云，且未在数据简介中做出说明，使用户无法查询、理解和利用。

上海在此项排名居于所有在评地区的前 10%

● 无问题 API

无无效 API：无效 API 是指存在于平台之上，但却无法调取的 API，主要包括以下三种情况：一是提交 API 接口调用申请后未收到任何反馈；二是 API 接口调用说明不清，无法调用；三是调用申请审核通过且按照调用说明正确操作后，仍无法获取数据。

上海在此项排名居于所有在评地区的后 30%。

无重复 API：重复 API 是指平台上出现标题相同或极其相似，实际调取到的数据集完全相同、所属主题也相同的 API 接口。

上海在此项排名居于所有在评地区的后 30%。

3.3 数据标准

(1) 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。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。

上海在此项排名居于所有在评地区的前 10%。

(2) 技术开放性

● 可机读格式：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。

上海在此项排名居于所有在评地区的前 10%。

● 非专属格式：是指可下载数据集应以开放的、非专属的格式提供，任何数据提供主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。

上海在此项排名居于所有在评地区的前 40%。

● RDF 格式：采用 RDF 格式发布数据集。

上海在此项排名居于所有在评地区的后 30%。

● API 开放性：是指申请与调用 API 接口的难易程度。

上海在此项排名居于所有在评地区的前 25%。

(3) 开放类型：平台对数据集明确标示了不同的开放类型，例如无条件开放、有条件开放和不予开放等。

上海在此项排名居于所有在评地区的前 10%。

(4) 描述说明：对数据集的描述与说明。

● 元数据提供：对 12 个基本元数据的提供情况。

上海在此项排名居于所有在评地区的前 10%。

● API 说明：包括基本调用说明与使用操作指南两种形式。基本调用说明是指提供了调用接口的基本说明，包括请求地址、请求参数与返回参数等信息。使用操作指南是指提供了详细的 API 接口使用方法与流程。

上海在此项排名居于所有在评地区的前 10%。

3.4 数据覆盖面

(1) 主题覆盖面：14 个基本主题在平台上的开放情况。

上海在此项排名居于所有在评地区的前 25%。

(2) 部门覆盖面：反映了各个政府部门开放数据的参与程度。

上海在此项排名居于所有在评地区的前 25%。

(3) 常见数据集覆盖面：14 个常见数据集在平台上的开放情况。

上海在此项排名居于所有在评地区的前 25%。

3.5 数据持续性

(1) 持续增长：平台从 2017 第二季度到 2019 年第一季度这一时间段内每季度新增数据集的情况。

上海在此项排名居于所有在评地区的前 10%。

(2) 动态更新：平台在 2019 年 1 月至 2019 年 3 月期间更新的数据集数量。

上海在此项排名居于所有在评地区的前 25%。

(3) 历史存档：平台留存了历史上不同时间发布的多个批次的数据，使用户可获取和利用。

上海在此项排名居于所有在评地区的前 10%。

3.6 建议

根据数据层各项指标上，上海在全国所处的位置，提出以下建议：

(1) 继续提升数据集总数，重点增加数据集的容量。

(2) 基于用户需求开放高容量、高需求的数据集，提高平台优质数据集与优质 API 接口的数量。减少平台上出现的部分碎片化与低容量数据集，减少无效 API 和重复 API 接口的数量。

(3) 提升平台上非专属格式数据集的比例，为用户提供 RDF 格式的数据。保持元数据丰富性，提供多种开放类型的数据集，包括有条件开放和不予开放，促进数据的分级分类开放。

(4) 进一步提升主题覆盖与常见数据集覆盖面。

(5) 保障平台日常运维，提升存量数据的更新频率。

4. 利用层

上海在利用层各项指标上的得分如图 5 所示。

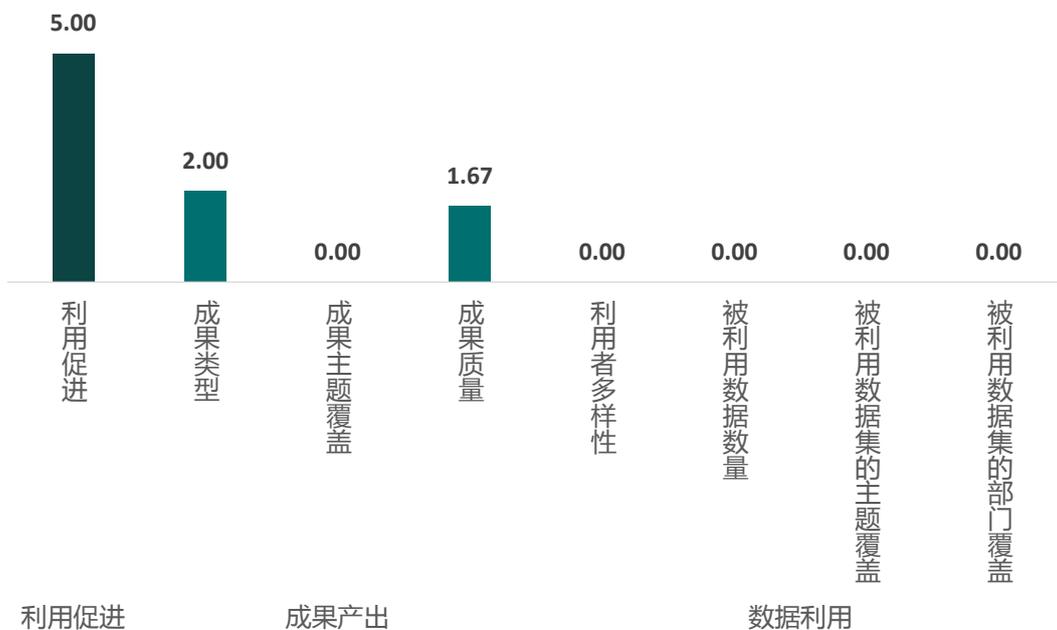


图 5 上海利用层各级指标得分

4.1 利用促进

利用促进是指政府为了推动开放数据的社会化利用而组织的各类活动。组织开放数据创新利用比赛是一种常见的利用促进活动。

上海在此项排名居于所有在评地区的前 10%。

4.2 成果产出

成果产出是指社会利用开放数据所产生的利用成果的类型、主题覆盖和质量。

(1) 成果类型：平台上展示的服务应用、创新方案、传播产品和研究成果等类型的利用成果。

●服务应用：有效的服务应用是指平台上展示的同时满足以下标准的应用：由数据利用者而非政府部门自身开发利用所产生；确实利用了开放的政府数据；能正常获取；能正常使用。

上海在此项排名居于所有在评地区的后 30%。

●创新方案：平台上展示的开放数据创新利用比赛参赛者产生的创新方案。

上海在此项排名居于所有在评地区的前 10%。

(2) 成果主题覆盖：平台上有效应用的主题覆盖面。

上海在此项排名居于所有在评地区的后 30%。

(3) 成果质量：平台上展示的开放数据利用成果的质量。

●优质服务应用：优质服务应用需要同时满足以下 4 个条件：符合有效服务应用的标准、利用的核心数据集为政府开放数据、利用了大容量的政府开放数据、除利用政府开放数据外，还融合利用了其他来源的数据。

上海在此项排名居于所有在评地区的后 30%。

●无问题成果：问题成果包括政府自身开发的成果、无效成果、数据来源不明成果和数据关联缺失的成果。

无政府自身开发成果：政府自身开发成果是指平台上展示的是由政府自身或委托第三方开发的政务服务类应用。

上海在此项排名居于所有在评地区的前 10%。

无无效成果：无效成果是指平台上展示了无法搜索到、能搜索到但无法下载或能下载但无法正常使用的服务应用。

上海在此项排名居于所有在评地区的后 30%。

无数据来源不明：数据来源不明是指平台上展示的利用成果没有标示数据来源。

上海在此项排名居于所有在评地区的后 30%。

无数据关联缺失：数据关联缺失是指利用成果虽然标注了所利用的开放数据集，但这些数据集在平台上链接不到、搜索不到或数据不完整。

上海在此项排名居于所有在评地区的后 30%。

4.3 数据利用

数据利用是指数据利用者为开放数据的利用方式和过程。

(1) 利用者多样性：平台上开放政府数据利用者类型的多样性。

上海在此项排名居于所有在评地区的后 30%。

(2) 被利用数据数量：平台上有效应用所使用的数据数量。

● 被利用数据集数量：平台上有效应用所使用的开放数据集的数量。

上海在此项排名居于所有在评地区的后 30%。

● 被利用数据集容量：平台上有效应用所使用的开放数据集的容量。

上海在此项排名居于所有在评地区的后 30%。

(3) 被利用数据集的主题覆盖：平台上有效应用所使用的开放数据集涵盖的主题，即这些应用利用了哪些主题下的开放数据。

上海在此项排名居于所有在评地区的后 30%。

(4) 被利用数据集的部门覆盖：平台上有效应用所使用的开放数据集覆盖到的部门，即这些应用利用了来自哪些政府部门的开放数据。

上海在此项排名居于所有在评地区的后 30%。

4.4 建议：

根据利用层各项指标上，上海在全国所处的位置，提出以下建议：

(1) 详细排查平台上展示的问题成果，重点剔除在平台上展示的无效成果、数据来源不明成果和数据关联缺失的成果。

(2) 从数据集数量和数据集容量两个方面提升被利用数据数量。

(3) 增加有效服务应用和优质服务应用数量。

(4) 鼓励多元利用者对政府数据进行利用。

(5) 提升被利用数据集的主题覆盖面和被利用数据集的部门覆盖面。

(6) 提高成果的主题覆盖面。

欲下载报告全文或定制地方指数分析报告可访问中国开放数林指数网站 (ifopendata.fudan.edu.cn) 或复旦大学数字与移动治理实验室网站 (www.dmg.fudan.edu.cn)，或发送邮件到 dmglab@fudan.edu.cn。



● 中国开放数林指数网 ifopendata.fudan.edu.cn



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



上海市杨浦区邯郸路220号复旦大学智库楼404室
dmglab@fudan.edu.cn
● www.dmg.fudan.edu.cn



关注
复旦DMG
公众号

