



中国开放数林指数
CHINA OPEN DATA INDEX

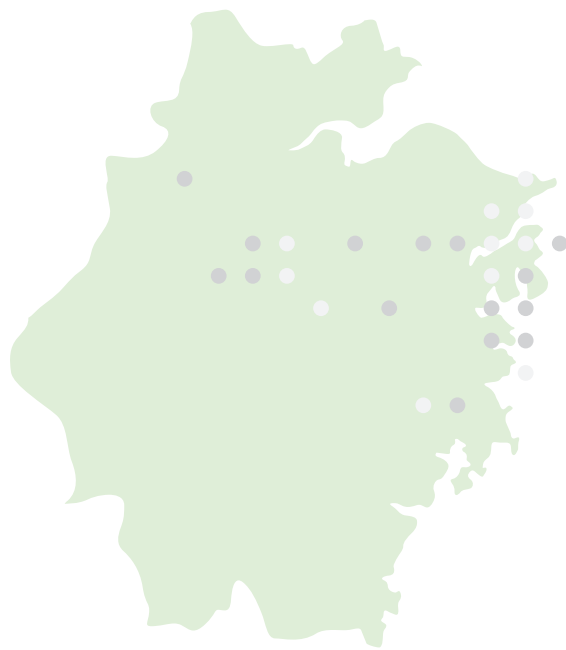
复旦智库报告

2020
上半年

地方开放数林指数分析报告

数据层 | 浙江

指数出品方 | 复旦大学数字与移动治理实验室
联合发布方 | 复旦大学 国家信息中心数字中国研究院





中国开放数据指数
CHINA OPEN DATA INDEX

· 2020 ·

CONTENTS

■ 评估方法	003
1.1 评估指标体系	003
1.2 数据采集分析方法	006
1.3 指数计算方法	006
■ 指数分析	008
2.1 数据数量	008
2.2 数据质量	009
2.3 数据规范	012
2.4 开放范围	017

请参照以下规范：

复旦大学数字与移动治理实验室. 地方开放数林指数分析报告 数据层_浙江 (2020 上半年) [R/OL]. (2020-07-22)[引用日期, 格式为 YYYY-MM-DD]. [http://ifopendata.fudan.edu.cn/static/papers/地方开放数林指数分析报告_数据层_浙江 \(2020 上半年\) .pdf](http://ifopendata.fudan.edu.cn/static/papers/地方开放数林指数分析报告_数据层_浙江(2020上半年).pdf).

评估方法

一、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 1-1、图 1-2)：

准备度是“数根”，是数据开放的基础，包括法规政策效力与内容、组织与实施、标准规范制定等三个一级指标。

平台层是“数干”，是数据开放的枢纽，包括数据发现、数据获取、成果提交展示、互动反馈、用户体验等五个一级指标。

数据层是“数叶”，是数据开放的核心，包括数据数量、数据质量、数据规范、开放范围等四个一级指标。

利用层是“数果”，是数据开放的成效，包括利用促进、有效成果数量、成果质量、利用多样性等四个一级指标。

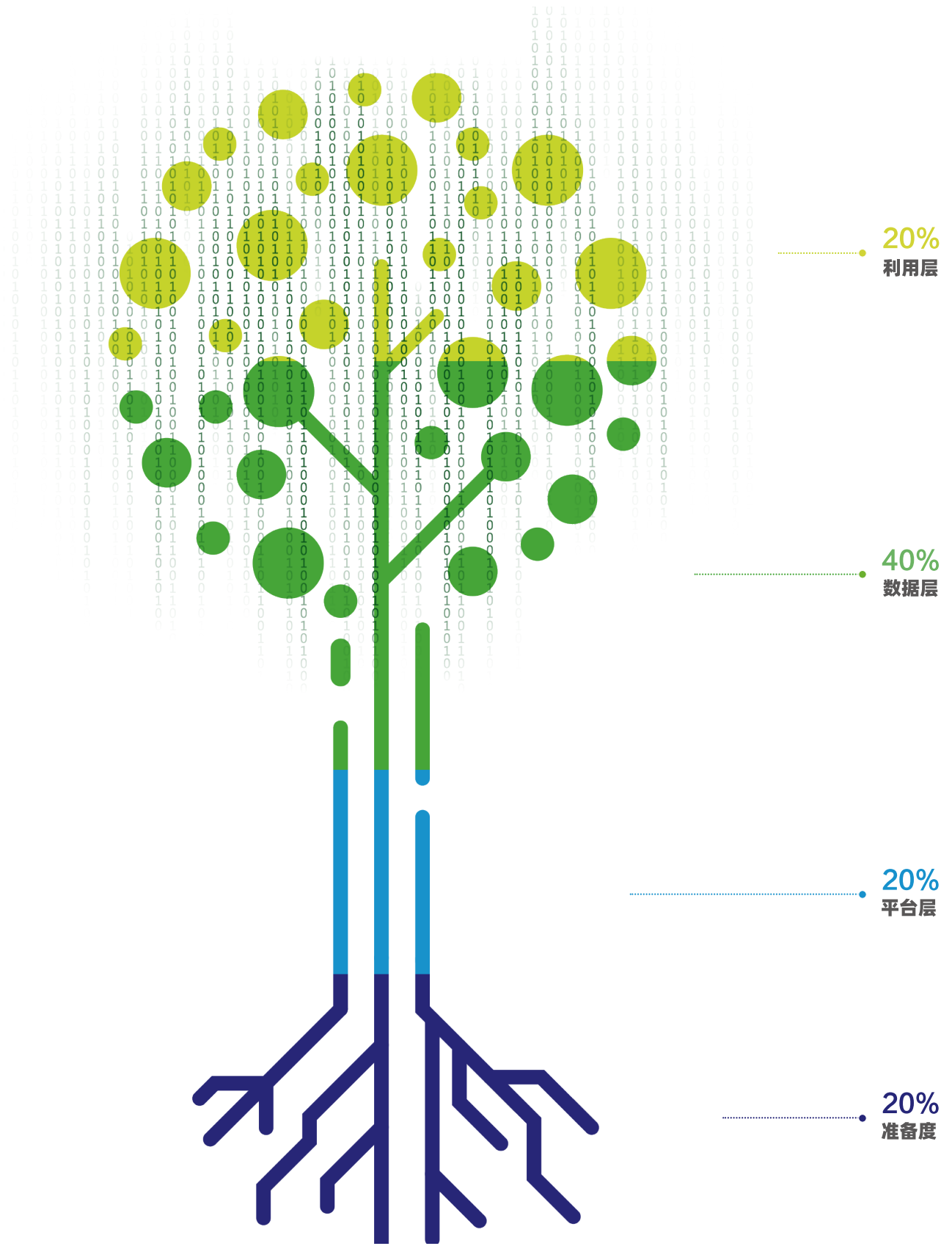


图 1-1 评估指标体系

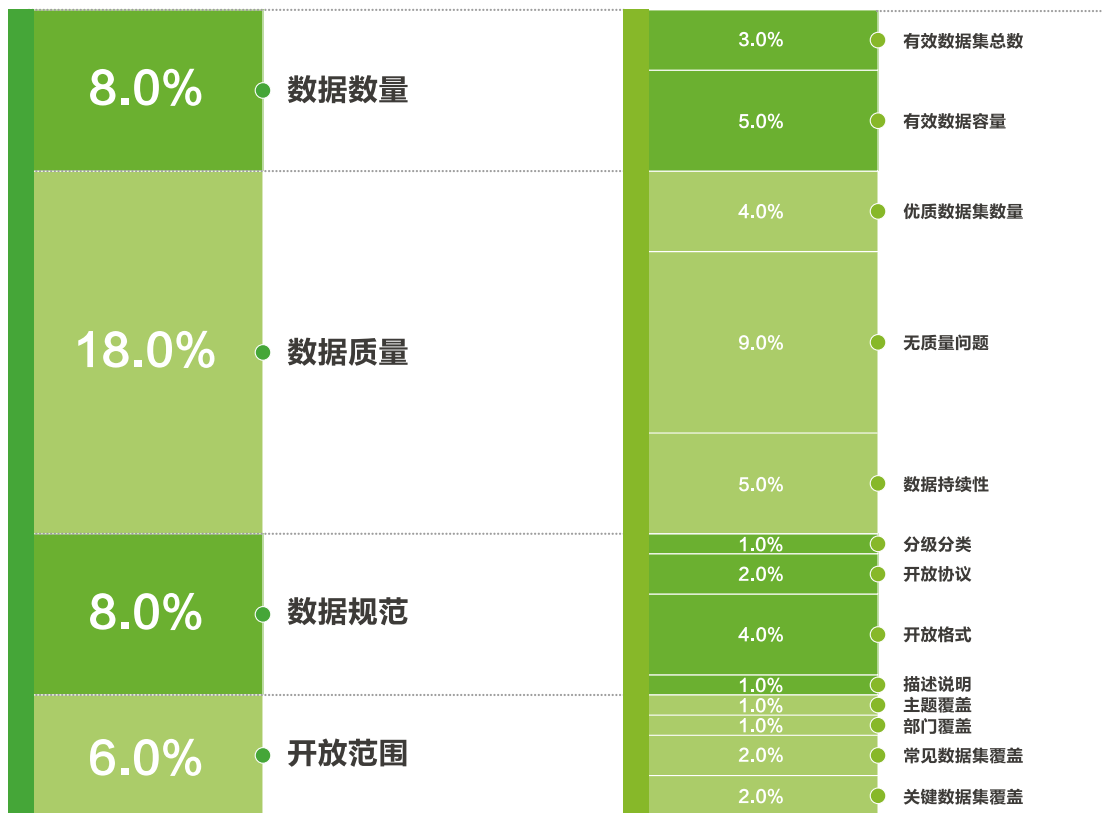


图 1-2 数据层评估指标

1.2 数据采集分析方法

数据层评估主要通过机器自动抓取浙江省数据开放平台上的数据，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 2020 年 4 月，对“动态更新”这一指标的评测时段为 2020 年 1 月 1 日至 2020 年 3 月 31 日。

1.3 指数计算方法

基于浙江省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 分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% 五个等级来展示浙江省在该项指标上所处的位置。

指数分析

二、指数分析

浙江在数据层各项指标上的得分如图 2 所示。

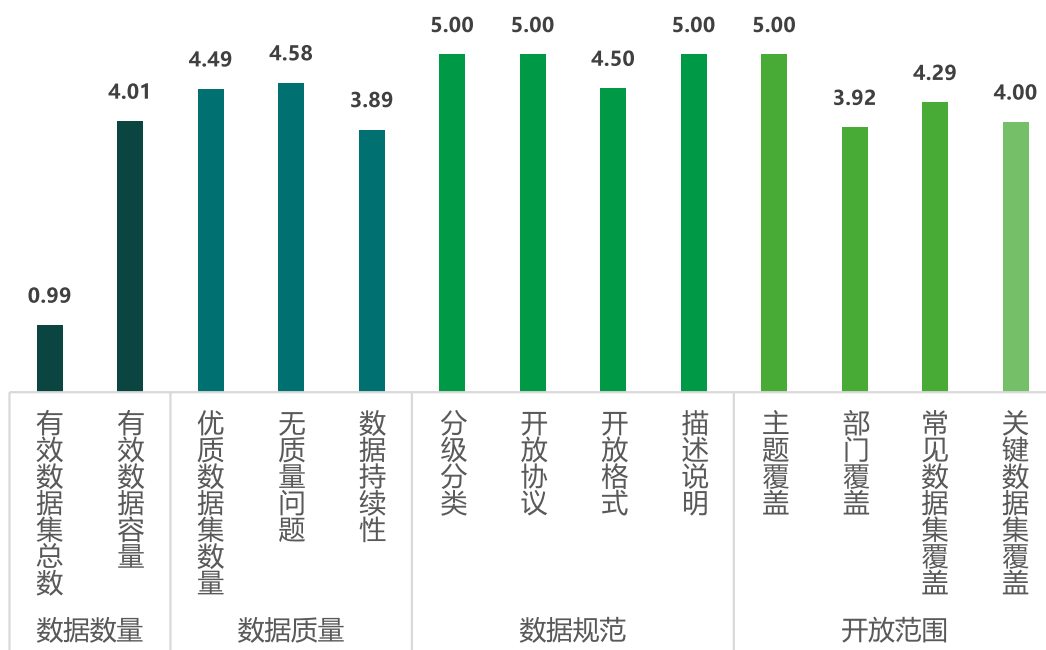


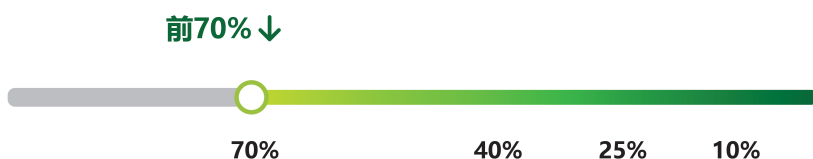
图 2 浙江在数据层各项指标上的得分情况

2.1 数据数量

数据数量是指平台上开放的有效数据集的数量和容量。

2.1.1 有效数据集总数

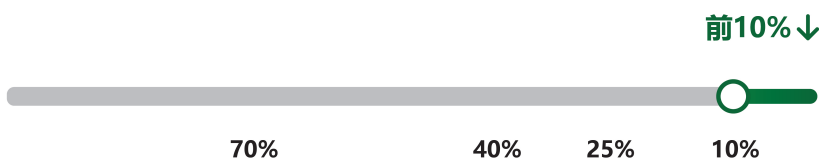
有效数据集总数是指平台上开放的真实有效的数据集的数量，不包括空白数据集、虚假数据集和重复数据集。浙江目前在该项指标上的排名居于在评地区的前 70%。



■ 浙江在该项指标上在全国所处的位置

2.1.2 有效数据容量

有效数据容量是指将一个平台上可下载的、结构化的有效数据集的字段数（列数）乘以条数（行数）后得出的数据量。浙江目前在该项指标上的排名居于在评地区的前 10%。



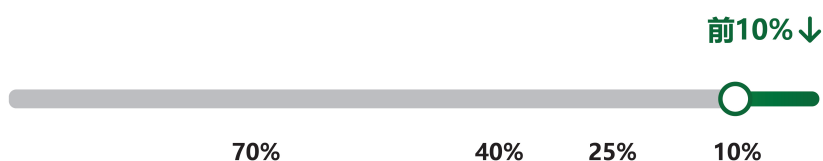
■ 浙江在该项指标上在全国所处的位置

2.2 数据质量

数据质量是指平台上开放的数据集的完整性、颗粒度、时效性和可获取性。

2.2.1 优质数据集数量

优质数据集数量是指平台上数据容量大、社会需求高的数据集的数量。报告对在评地区平台上所有可下载的数据集按照数据容量进行排序，在数据容量相同的情况下再按照下载量排序，最终选出排名居于前 1% 的数据集作为优质数据集。浙江目前在该项指标上的排名居于在评地区的前 10%。



■ 浙江在该项指标上在全国所处的位置

浙江省平台有三个优质数据集进入在评地区优质数据集的前十名（如表 1）。

■ 表 1 浙江平台上开放的优质数据集

序号	数据集名称	地方	数据容量	条数	字段数
1	单位未履行生效裁判信息	浙江	18136621	788547	23
2	道路运输证信息	浙江	13210494	600477	22
3	公积金个人贷款信息	浙江	9511282	864662	11

2.2.2 无质量问题

无质量问题是指平台上存在高缺失、碎片化、低容量、生硬格式转化、限制型 API、标题缺失或不清等质量问题的数据集较少或没有。其中，高缺失数据集是指数据集中有 60% 以上的空缺数据；碎片化数据集是指按照时间、行政区划、政府部门等被人为分割的数据集；低容量数据集是指因数据量本身稀少或颗粒度过大等原因造成的数据条数在三行或三行以内的数据集；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。浙江目前在该项指标上的排名居于在评地区的前 25%。



■ 浙江在该项指标上在全国所处的位置

2.2.3 数据持续性

数据持续性是指平台能够持续增加数据集，更新已开放数据集，并留存历史数据集。浙江目前在该项指标上的排名居于在评地区的前 10%。



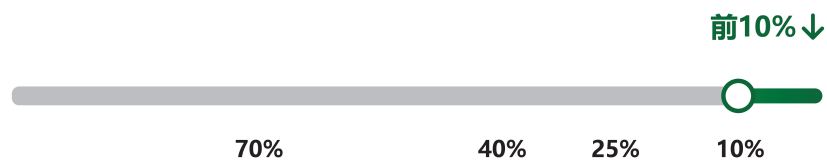
■ 浙江在该项指标上在全国所处的位置

2.3 数据规范

数据规范是指平台上的开放数据集在分级分类、开放协议、格式标准、描述说明等方面的规范和完善程度。

2.3.1 分级分类

分级分类是指平台根据不同的安全保护要求对数据集标示了不同等级的开放类型，浙江目前在该项指标上的排名居于在评地区的前 10%。



■ 浙江在该项指标上在全国所处的位置

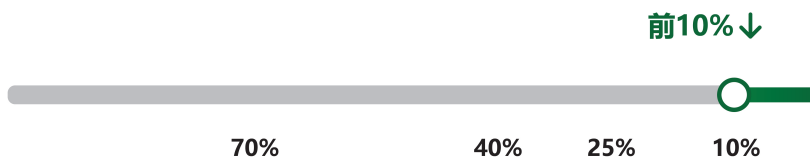
浙江平台为数据集标注了向社会开放的类型，包括无条件开放与受限开放（如图 3）。

信息资源名称	信息资源代码	信息资源摘要	数据领域	是否向社会开放	开放条件
优秀教育工作者信息	30701604000000/869	优秀教育工作者信息,包含姓名,性别,	教育文化	受限开放	开放网站
商品住宅成交备案均价信息	30701604000000/567	商品住宅成交备案均价信息,包含1-X	其他	受限开放	开放网站
湖州市体育项目成绩破、超(平)汇	30701604000000/901	湖州市体育项目成绩破、超(平)汇	教育文化	受限开放	开放网站
食用农产品生产主体信用评价信息	30701600003034/2081	食用农产品生产主体信用评价信息,	信用服务	受限开放	开放网站
地震监测台站信息	30701600003222/520	地震监测台站,包含场所地责任人	公共安全	受限开放	开放网站
建筑起重机械设备安装告知信息	30701603000000/3878	建筑起重机械设备安装告知,包含是	城建住房,安全生产,市场监管	受限开放	开放网站
优秀团员信息	30701604000000/940	优秀团员信息,包含表彰时间,表彰文	教育文化	受限开放	开放网站
基金会登记信息	30701600000000/94164	基金会登记信息,包含业务主管单位,	机构团体	受限开放	开放网站
中介机构名单信息	30701600000000/30679	中介机构名单信息,包含证书编号,联	生活服务	受限开放	开放网站
永嘉县果茶生产大户名单信息	30701600000000/30239	永嘉县果茶生产大户名单信息,包含	工业农业	受限开放	开放网站
农经站农民专业合作社清单信息	30701600000000/30238	农经站农民专业合作社清单信息,包含	工业农业	受限开放	开放网站
养老机构信息	30701600000000/94166	养老机构信息,包含地址,联系电话,发	其他	受限开放	开放网站
温州市城市社区基本情况统计信息	30701600000000/30456	温州市城市社区基本情况统计信息,	城建住房	受限开放	开放网站
事业单位信息	30701600000000/30443	事业单位基本情况简表信息,包含年	生活服务	受限开放	开放网站
规上工业亩均税收信息	30701600003020/43	规上工业亩均税收,包含地区名称,地	其他	无条件开放	开放网站
规上工业亩均增加值信息	30701600003020/42	规上工业亩均增加值,包含地区名称,	其他	无条件开放	开放网站
新增工业机器人台数	30701600003020/296	新增工业机器人台数,包含主键,地区	市场监管	无条件开放	开放网站
淘汰落后产能涉及企业统计信息	307016001003084/206	淘汰落后产能涉及企业数,包含地	科技创新	无条件开放	开放网站

■ 图 3 浙江平台开放数据目录中标注的开放类型

2.3.2 开放协议

开放协议是指平台为数据集提供了有关其开放主体和利用主体的权利和义务的协议，浙江目前在该项指标上的排名居于在评地区的前 10%。





■ 浙江在该项指标上在全国所处的位置

浙江平台在用户获取数据时提供了数据开放授权许可使用协议（如图 4）。

浙江省数据开放平台数据开放授权许可使用协议

发布日期：2019-02-27 13:37 浏览次数：1

分享到： 

在您访问或使用浙江省数据开放平台前，请认真阅读本协议。

如使用本平台，您必须完全接受本协议中所包含的条款、条件和平台即时刊登的通告，并且遵守中华人民共和国关于互联网的相关法律、规定与规则。

本平台有权不经提前通知随时修改本协议。本协议一经变动，将在相关页面提示修改内容；用户如果不同意本协议的修改，可以放弃访问或使用浙江省数据开放平台提供的任何数据、产品和服务；如果用户在协议变动后继续访问或使用浙江省数据开放平台，则视为已经接受协议的变动。

一、用户权利和义务

1. 现阶段，用户有权免费获取本平台所提供的所有公共数据资源，享有数据资源的非排他使用权。用户可以自由地复制、发布、利用、传播和分享信息，不受歧视。
2. 用户在使用本平台数据资源所产生的成果中应注明公共数据资源来源为“浙江省数据开放平台”，及从平台下载的时间。
3. 用户基于本平台公共数据资源开发的各类应用和服务，应遵守有关法律、法规和规章的规定，不得用于任何可能危害国家安全和公共公共利益、侵犯商业秘密和个人隐私或其他不正当的用途。因与第三方产生纠纷或者违反法律法规等造成的后果，由用户自行承担。
4. 用户及时将应用情况通过网站进行备案，并积极配合有关的用户需求调查和数据资源调查。
5. 用户基于本平台公共数据资源进行开发利用时，应清楚地表明所做的任何数据分析或应用是个人或公司的行为，不得歪曲篡改从本平台获取的公共数据资源。
6. 用户在基于本平台数据资源的应用程序或互联网服务中应清楚地声明源于平台的相关数据受浙江省数据开放平台相关协议、条款约束。以上义务同样需要得到您的再授权用户同意并认可。
7. 如果本平台数据服务内容由于法律法规或政策调整等原因进行调整，用户应立即在相关应用程序或互联网服务中作相应调整。
8. 用户应遵守中华人民共和国有关互联网管理的相关法律法规和规章，以及所有与本平台网络服务有关的网络协议、规定和程序。禁止对本平台进行技术性破坏，不得干扰或妨害浙江省数据开放平台提供正常的互联网服务。用户发现任何非法使用用户帐号或安全漏洞的情况，应当立即告知本平台运营管理方。
9. 用户义务是本许可的重要条件，如未遵守以上要求，则根据本许可授予的权利将自动终止，用户应立即删除相关数据并停止服务。

二、平台权利及免责声明

1. 因国家法律法规和政策调整，或因多条数据关联后可能造成隐私泄露及其它原因不宜继续公开的数据，本平台有权随时将其下线，用户不得以任何理由继续保存或使用该数据，否则造成的后果由您自行承担。
2. 本平台所有数据资源均为公共管理和服务部门在履职过程中采集和获取，由于未采集新数据或其他原因可能造成数据质量波动，及更新推迟或中止。本平台对所提供数据的完整性、准确性、及时性和服务连续性不作任何承诺或担保，并且不对因使用本平台所提供的数据或服务可能造成的任何损失承担责任。用户应自行判断是否继续使用本平台提供的数据或服务。
3. 本平台不对由于操作或传播的延迟、通讯失败、互联网接入设备的困难，硬件或软件故障而造成的直接或间接损失承担责任。
4. 任何由于黑客攻击、计算机病毒侵入或发作、政府管制而造成的暂时性关闭等影响网络正常经营及其他不可抗力原因而造成的资料泄露、丢失、被盗用或被篡改等，本平台将采取必要措施尽力减少用户的损失，但本平台对此不承担任何责任。
5. 浙江省数据开放平台包含到其他网站的链接，这些网站内容和隐私策略可能和我们不同。本平台不对由于这些网站内容和隐私策略造成的后果承担责任。

三、隐私保护声明

本平台尊重并保护所有用户的个人隐私权，未经用户许可或根据相关法律、法规或规章规定，以及维护浙江省数据开放平台合法权益之外，本平台不会将用户个人信息提供给任何第三方，但是会根据工作需要与政府有关部门共同使用。

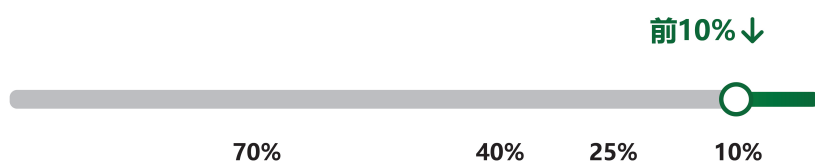
四、条款说明

本协议条款的最终解释权归浙江省大数据发展管理局所有。

图 4 浙江平台的数据开放授权许可使用协议

2.3.3 开放格式

开放格式是指平台上开放数据集的格式标准，包括可机读格式、非专属格式、RDF 格式和 API 调用标准。浙江目前在该项指标上的排名居于在评地区的前 10%。



■ 浙江在该项指标上在全国所处的位置

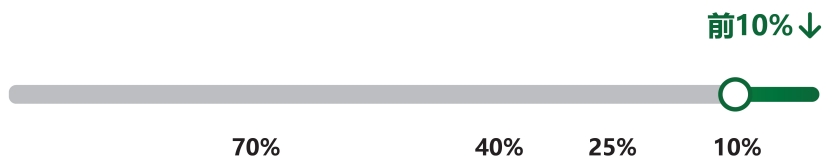
浙江为平台上的数据集提供了可机读格式、非专属格式与 RDF 格式（如图 5）。

行业分类:	暂无	数据预览:	预览 可视化
数据分级:	登录开放	数据接口:	API
开放条件:	开放网站	数据范围:	省市县
更新日期:	2020-08-16	发布日期:	2019-06-20
数据量:	268	数据图谱:	数据图谱
数据下载:	XLS CSV XML JSON RDF		

■ 图 5 浙江平台提供的数据集下载格式

2.3.4 描述说明

描述说明是指对数据集的描述与说明，包括元数据提供与 API 说明。浙江目前在该项指标上的排名居于在评地区的前 10%。



■ 浙江在该项指标上在全国所处的位置

浙江平台为数据集提供了完整的元数据与 API 说明（如图 6）。

拖欠工资黑名单信息 ★★★★★

摘要:	拖欠工资黑名单,包含失信黑名单项目名称,发布机关名称,法人和其他组织名称,统一社会信用代码,发布时间等信息		
资源代码:	307016001003085/152		
标签:	工资、信息、拖欠、黑名单	更新周期:	不定期
资源格式:	数据库-oracle	数据源单位:	省人社保厅
数据源单位地址:	杭州市省府路8号省府大院2号楼	联系方式:	0571-87057390
数据领域:	社保就业	访问/下载次数:	17869/1375
系统名称:	浙江省劳动保障监察信息系统	评分/评价次数:	5/5
行业分类:	暂无	数据预览:	预览 可视化
数据分级:	登录开放	数据接口:	API
开放条件:	开放网站	数据范围:	省市县
更新日期:	2020-08-16	发布日期:	2019-06-20
数据量:	268	数据图谱:	数据图谱
数据下载:	XLS CSV XML JSON RDF		

拖欠工资黑名单信息 ★★★★★ ★ 收藏 [纠错](#)

摘要:	拖欠工资黑名单,包含失信黑名单项目名称,发布机关名称,法人和其他组织名称,统一社会信用代码,发布时间等信息		
数据领域	社保就业	申请接口	申请
更新周期	不定期	发布时间	2019-06-20
数据提供单位	省人社保厅	更新时间	2020-08-16
访问次数	940	调用次数	22542
接口预览	预览	数据量	268
数据预览:	预览		

[总数接口](#) | [分页接口](#) | [分页返回结果](#) | [数据更新查询接口](#)

接口地址: interfaces/cata_5994/get_total.do

支持格式: JSON

请求方式: GET/POST

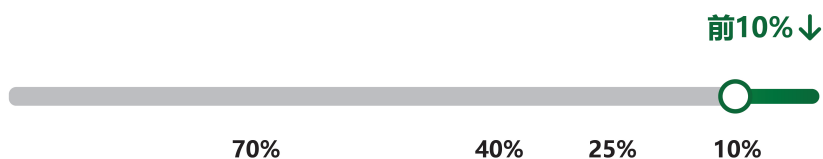
图 6 浙江平台的元数据与 API 接口说明

2.4 开放范围

开放范围是指平台上开放的数据集在数据主题、参与部门、常见数据集及关键数据集方面的覆盖程度。

2.4.1 主题覆盖

主题覆盖是指平台上开放的数据集在经贸工商、交通出行、卫生健康、教育科技、社保就业、资源环境等 14 个基本主题上的覆盖程度，浙江目前在该项指标上的排名居于在评地区的前 10%。



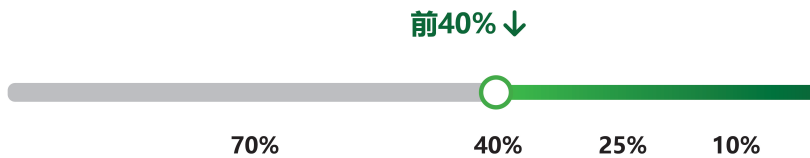
■ 浙江在该项指标上在全国所处的位置

浙江平台的数据覆盖了所有基本主题（如表 2）。

经贸工商	✓	资源环境	✓
交通出行	✓	城建住房	✓
机构团体	✓	公共安全	✓
文化休闲	✓	农业农村	✓
卫生健康	✓	社保就业	✓
教育科技	✓	财税金融	✓
社会民生	✓	信用服务	✓

2.4.2 部门覆盖

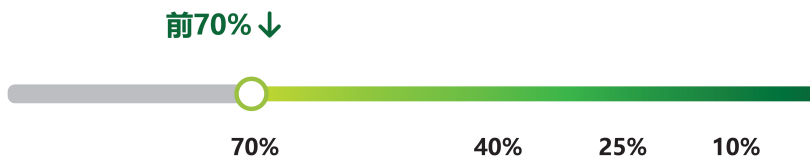
部门覆盖是指平台上开放的数据集所覆盖到的政府部门的比例，浙江目前在该项指标上的排名居于在评地区的前 40%。



■ 浙江在该项指标上在全国所处的位置

2.4.3 常见数据集覆盖

常见数据集覆盖是指平台上开放的数据集在常见数据集上的覆盖程度。报告对在评地区平台上开放的所有数据集的名称进行了文本分析，梳理出十四类各地平台上常见的开放数据集。浙江目前在该项指标上的排名居于在评地区的前 70%。



■ 浙江在该项指标上在全国所处的位置

2.4.4 关键数据集覆盖

关键数据集是指平台上开放的数据集在关键数据集上的覆盖程度。报告基于《中共中央 国务院关于构建更加完善的要素市场化配置体制机制的意见》和地方法规政策要求重点和优先开放的数据领域，结合各地平台上已开放的优质数据集清单，梳理出五项关键数据集。浙江目前在该项指标上的排名居于在评地区的前 10%。



■ 浙江在该项指标上在全国所处的位置

浙江平台提供了除疫情防控类数据以外的关键数据集（如表 3）。

企业注册登记类数据	✓
公交车辆位置类数据	✓
道路运输从业资格证/经营许可证类数据	✓
气象预报预警类数据	✓
疫情防控类数据	

定制地方开放数林指数分析报告可在公众号“复旦 DMG”下方留言或者回复关键词“地方报告”，也可发送邮件到 dmglab@fudan.edu.cn

联合发布方



合作单位

冥睿（上海）信息科技有限公司

汇纳科技股份有限公司

DataMate 数据工作室

晴禾（南京）文化有限公司

浪潮集团

复旦发展研究院

中山大学数字治理研究中心

上海市静安区人力资源和社会保障局



中国开放数据指数
CHINA OPEN DATA INDEX

· 2020 ·



中国开放数林指数网 ifopendata.fudan.edu.cn

合作方：

冥睿（上海）信息科技有限公司
汇纳科技股份有限公司
DataMate数据工作室
晴禾（南京）文化有限公司
浪潮集团
复旦发展研究院
中山大学数字治理研究中心
上海市静安区人力资源和社会保障局



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