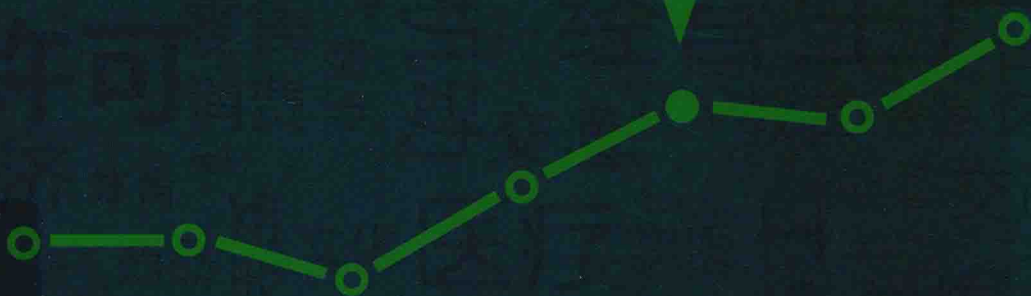


# 开放的 数据林

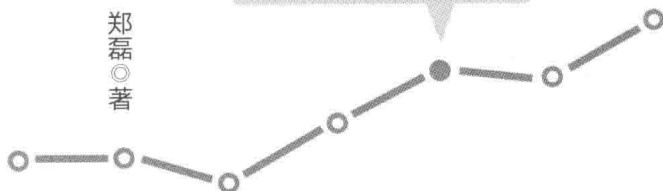
政府数据开放的  
中国故事

郑磊◎著



开放的  
数林  
政府数据开放的  
中国故事

郑磊◎著



## 图书在版编目(CIP)数据

开放的数林:政府数据开放的中国故事/郑磊著

—上海:上海人民出版社,2018

ISBN 978-7-208-15349-3

I. ①开… II. ①郑… III. ①电子政务-信息管理-研究-中国 IV. ①D630.1

中国版本图书馆 CIP 数据核字(2018)第 161683 号

责任编辑 秦 堃 夏红梅

封面设计 零创意文化

## 开放的数林:政府数据开放的中国故事

郑 磊 著

出 版 上海人民出版社  
(200001 上海福建中路 193 号)  
发 行 上海人民出版社发行中心  
印 刷 上海商务联西印刷有限公司  
开 本 720×1000 1/16  
印 张 15.5  
插 页 2  
字 数 265,000  
版 次 2018 年 8 月第 1 版  
印 次 2018 年 10 月第 2 次印刷  
ISBN 978-7-208-15349-3/D·3256  
定 价 56.00 元

# 序

## 开放数据 蔚然成林

2006年10月中旬,我和郑磊老师第一次见面,那时他作为美国纽约州立大学的博士生陪同导师 Sharon Dawes 教授一行来北京参加一个国际电子政务研讨会。自此以后的12年时间里,我们都投身到中国数字治理和大数据工作中,所不同的是,他在象牙塔里教书育人做研究,我在政府事业单位一线做些具体工作,他是学术圈里最接近一线工作、最有实战经验的少数青年学者之一,我自诩是一线工作者中比较有学术情怀的。

12年里,我们共同见证了中国政府信息化的飞速发展,特别是最近几年,大数据上升为国家战略后,我们在政府数据开放这个领域有了更多交集,在研究和实践的相互碰撞中,我们意识到,在这个全球数字化转型的重要历史时期,必须把握大数据发展机遇,跟上新时代的步伐。同时,我们更意识到,大数据在中国的发展路径是什么、政策框架是什么、制约因素是什么、风险边界是什么,等等,这些关键问题都需要逐一回答。

从欧美发达国家的经验看,推动政府数据开放是大数据战略中必不可少的关键环节,是政府治理创新的重要手段,更是带动数字经济从而促进高质量发展的重要抓手,从面向中国未来改革创新的角度看,可以说怎么强调这一工作的重要性都不为过。从当前我国政府数据开放的进展来看,总体还处在起步阶段,从认识到实践还存在很多模糊的甚至是错误的东西。在认识层面,到底什么是政府数据开放?从开放数据到带动大数据产业实现突破式发展的传递机制是什么?数据开放在哪些环节可以促进政府治理创新?对这些问题普遍缺乏基本的认知。在实践层面,统计数据、政策文件上网算不算数据开放?政府数据开放所有工作一定都要由政府来做吗?提供数据公共服务与生产公共服务在什么条件下可以分离?这些问题从各地五花八门的做法上都能够清楚的反映出来。我国政府数据开放工作的发展已经对全面系统梳理政府数据开放的理论和实践提出非常紧迫的需要,因此,郑磊老师这部专著的出版就显得尤为及时和必要,书中对基本概

念、认识和做法的深入分析，不仅对于学术界有很高价值，对于指导各级政府从事数据开放的工作者而言，具有极高的“手册”意义。

近几个月，郑磊老师的团队积极参与国家公共数据开放平台建设的相关研究咨询工作，为我们提出了很多有价值的意见和建议。在讨论中，我们形成了一个共识，从发展历史背景和问题导向看，中国的政府数据开放与西方有所不同，政府数据开放不仅仅是政府治理创新和经济转型发展的助推剂，更重要的是对于有着“差不多先生”这顶帽子的中国人而言，政府数据开放更是一个数据文化缔造的崭新过程，因此，政府数据开放就不仅仅是技术、产业和管理创新问题，还具有文化创新的划时代意义。通过数据开放工作推动打造全社会尊重数据、保护数据、善用数据的文化氛围，使数据科学和数据产业成为青年人就学就业的向往，使大数据更好的服务于广大人民群众对美好生活的追求，让人的发展更加自由、政府运作更加透明、社会关系更加友善、市场行为更加诚信、科学研究更加高效。这是一条漫长的路，需要有“功成不必在我”“功成必定有我”的精神和意志，小步快走，行稳致远。

郑磊老师是我非常敬重的青年学者，不仅仅是因为他的国际化视野、严谨的治学精神和开放的互联网思维，更为重要的是在他身上有当下中国学术界少有的师道人格，他曾经为了学生发表论文署名的公正性与杂志社据理力争，他在学院新生入学和老生毕业典礼上的演讲曾经在网上传被大家争相传阅，在他的演讲中传递着朴实无华的正能量，他还积极投身公益慈善事业，为西部贫困地区学生提供无私资助。

中国有句古话叫“种豆得豆、种瓜得瓜”，在最近的一次大型数据开放报告发布活动中郑老师的团队给每个人发了两颗绿植种子，一颗上印着“开放数据”，另一颗上印着“蔚然成林”，我想这本专著正是为我国政府数据开放种下了一颗种子，终有一天，在所有数据开放工作者的共同耕耘、浇灌、呵护下，会豆熟瓜落，会大树参天。

于施洋

国家信息中心大数据发展部主任、数字中国研究院院长

2018年7月于北京

# 前 言

政府在治理社会服务公众的过程中，需要采集和生成大量数据，然后根据这些数据作出更科学的决策，进行更高效的管理，提供更精准的服务。更重要的是，政府所掌握的这些数据不仅对政府有用，也和老百姓的生活息息相关，涉及天气、食物、交通、上学、就业、买房、结婚、生娃、游玩、看病、养老等方方面面，可谓无所不包。

那么问题来了，这些政府数据除了政府自己可以拿来用，是不是也可以开放给社会，让人民来使用呢？答案是肯定的，因为这些数据本来就是属于人民的，是用人民的钱采集的，是为了人民而采集的，自然也应该开放给人民使用。当然也有例外，就是有些政府数据开放出来可能会损害人民的利益，比如涉及国家安全、商业机密和个人隐私的数据，这些数据不仅不能开放，还需要进行严格保护。

政府是人民的政府，政府的数据归根结底也是人民的数据，把政府数据开放出来，赋予人民使用这些数据的权利，可以让人民的生活变得更加美好。企业和社会公众通过对这些数据进行分析，或利用这些数据开发出各种各样的应用，可以让人民出行更便捷，吃饭更放心，环境更安全，居住更舒适，身体更健康，工作更顺利，创业更容易，作各种各样的决定时也更有依据。

因此，本书就来谈一谈政府数据开放这件大事，并讨论这样一些具体问题：为什么政府应该开放数据？哪些政府数据可以开放？哪些不可以开放？怎样才是开放政府数据的正确姿势？政府开放数据会带来哪些好处？又可能带来哪些风险？政府数据开放这件事和哪些人、哪些组织有关？目前我国开放政府数据最多最好的地方是哪里？还存在哪些问题？为什么开放政府数据这件事这么难？怎样才能开放更多更好的政府数据？政府数据开放的未来会怎样？

在结构上，本书将依次讲述和分析中国政府数据开放的背景、目的、现状、困境、路径和未来挑战。其中，第一、第二、第三章是回顾了国内外政府数据开放的背景和历程，阐述其价值、风险和利益相关者，并厘清一些基本概念与原则，为之后展开政府数据开放的中国故事搭建舞台；第四、第五章是基于一手数据和资料，

描绘了我国政府数据开放的整体现状和典型案例，为后面将要展开的理论分析提供素材和依据；第六、第七章是对中国政府数据开放面对的困境与纠结进行了提炼总结，并进一步探讨中国政府数据开放的发展路径；第八章则是为数据开放未来可能面对的新挑战抛出一些问题，引发思考和讨论。

本书既引用了许多国内外学术文献和研究报告，更采用了大量来自中国本地的一手数据和访谈资料，力求在理论研究与实践案例、国际视野与本土视角、“深度有据”和“易读有趣”之间找到一个平衡点，希望能为我国政府数据开放的研究者、实践者和关注者，包括来自政府部门、研究机构、高等院校和产业界的读者们，提供一本集理论梳理、实践解读和路径指南于一体的参考读物。

目前国内有关政府数据开放的著作多停留于介绍国外实践、引进国外理论或直接翻译国外论著阶段。作为一名留学回国的研究者，我对中外差异有切身体会，深知国外政府数据开放的研究发现和政策建议并不能直接应用于中国实践。虽然数据开放的理念最初来自国外，但要在中国真正落地，就必须根植于中国的本土环境，学术界也要针对中国面对的现实问题开展研究，讲述和剖析政府数据开放的中国故事。我也相信，只有基于中国实践的研究发现和经验，才能真正与国际政府数据开放的研究和实践展开对话，贡献中国思路。

自2015年国务院印发《促进大数据发展行动纲要》和2017年中央全面深化改革领导小组通过《关于推进公共信息资源开放的若干意见》之后，我国越来越多的地方政府加入数据开放的行列中。截至2018年5月，我国已上线了近五十个地级以上的数据开放平台，其中又有半数以上是在最近的一年中推出的。2018年年底，我国还将建成国家公共数据开放平台，相信又会带动更多地方政府开放数据，中国政府数据开放已进入快车道，也为我们的研究提供了极好的研究环境和时机。

面对这一形势，我们团队对各地政府数据开放平台的现状与问题进行了系统评测和分析，并于2017年5月起连续两年在贵阳国际大数据产业博览会上发布《中国地方政府数据开放报告》和“中国开放数林指数”。同时，我们也受到一些中央部委和地方政府的邀请，深入参与政府数据开放的决策咨询、制度建设、规划制定、执行实施和考核测评的过程中，还参与策划和组织了上海开放数据创新应用大赛(SODA)，从而有机会对中国政府数据的开放和利用展开近距离观察和研究，了解到许多藏在表象之后的故事。在此期间，我们陆续在《电子政务》、《中国行政管理》、《公共行政评论》和《图书情报工作》等期刊上发表了十多篇政府数据开放方面的论文。

本书正是我们团队这五六年来研究工作的一个阶段性成果。我们研究中国

政府数据开放的过程,始终伴随着中国本地实践的发展,受到实践的启发,被实践不断推动,并对现实需求和真实问题作出回应。可以说,没有中国各地的实践,就不会有我们这些年的研究成果。同时,我们也很高兴地看到,我们的学术成果和研究报告逐渐被一些政府部门应用到实践中,助推其发展,实现了理论与实践的结合。

2018年正值我国改革开放40周年,改革再出发,开放再扩大,数据开放可以说是扩大开放在大数据领域的具体体现,也是政府部门在大数据时代背景下对改革开放40周年的最好纪念。本书能恰巧在这样一个重要的年份出版,既深感荣幸,又倍感压力。

最后想强调一下,本书只是一个阶段性的总结,并不是中国政府数据开放的最终成果,更不是标准答案,必然还有许多不成熟、不准确的地方。就此先抛出一个靶子,然后静待大家的砖块吧!

# 目 录

序 1

前言 1

## 第一章 没有数据开放就没有智能时代

——政府数据开放的前世今生 1

第一节 政府数据开放的起源与背景 1

一、封闭还是开放政府,这件事和数据有关 1

二、从开放源代码到开放数据 2

三、政府数据是智能时代的一种公共基础设施 3

第二节 全球政府数据开放:尚未成功,仍需努力 4

一、全球政府数据开放的兴起 4

二、国际合作加快政府数据开放步伐 6

三、全球政府数据开放的现状与挑战 7

第三节 中国政府数据开放:星星之火,可以燎原 9

## 第二章 政府开放数据以后会怎样?

——政府数据开放的价值、潜在风险和利益相关者 12

第一节 为什么要开放政府数据? 12

第二节 政府数据开放的收益与风险 18

一、经济、社会、政治和其他价值 18

二、政府数据开放有哪些潜在风险? 21

第三节 政府数据开放关系你我他 22

一、政府数据的提供者 22

二、政府数据的利用者 22

## 2 开放的数林:政府数据开放的中国故事

### 三、政府数据的用户和合作者 23

#### 第四节 数据开放是一个生态体系,政府是一个平台 23

## 第三章 真假数据开放

### ——什么是真正的政府数据开放? 27

#### 第一节 数据产品不等同于数据开放 27

##### 一、政府统计报告 27

##### 二、政府数据可视化 27

##### 三、政府数据查询服务 29

##### 四、政府数据应用 29

#### 第二节 真正的政府数据开放 30

#### 第三节 政府数据开放和其他概念的区别? 32

##### 一、开放数据、大数据和开放政府 32

##### 二、政府数据开放和政府信息公开 33

##### 三、开放数据和公开数据 34

##### 四、数据开放、数据共享和数据封闭 35

##### 五、数据开放和数据交易 36

##### 六、政府数据开放和公共数据开放 37

## 第四章 我国政府数据开放整体水平如何? 40

#### 第一节 我国政府数据开放在国际性评估中的位置 40

##### 一、从“开放数据晴雨表”看 40

##### 二、从“全球开放数据指数”看 42

#### 第二节 构建符合我国实际的政府数据开放评估指标体系与方法 43

##### 一、评估指标体系 43

##### 二、评估范围 53

##### 三、数据采集与分析方法 56

##### 四、指标计算方法 56

#### 第三节 我国地方政府数据开放概况 57

##### 一、上线时间 57

##### 二、地区分布 57

##### 三、行政层级 58

四、平台类型	59
第四节 我国地方政府数据开放的数量、质量、标准与价值	60
一、数据数量	60
二、数据质量	61
三、数据标准	63
四、数据覆盖面	70
五、数据持续性	75
第五节 我国地方政府数据开放平台的功能与体验	77
一、平台概览	77
二、平台导引	81
三、数据获取	87
四、工具提供	90
五、利用成果展示	94
六、互动交流	97
七、个性化整合	103
第六节 我国地方政府数据开放的基础与准备	105
一、法规与政策	105
二、领导力	115
三、组织保障	117
第七节 我国地方政府数据开放的现状与问题	120
一、指数分值与排名	120
二、综合分析	122
三、数据层分析	134
四、平台层分析	135
五、准备度分析	138
第五章 我国政府数据开放的先行者们进行了哪些探索?	141
第一节 先行先试,打造数据开放生态体系:上海市	141
一、率先开放:四大举措启动数据开放工作	141
二、持续推进:制定数据开放规划和工作计划	143
三、培育生态:组织数据创新大赛推动应用	145
四、规范管理:制定管理办法应对问题与风险	149

## 4 开放的数林:政府数据开放的中国故事

五、纵深发展:试点区级层面的数据开放	150
第二节 全速推进,后来居上引领全国:贵阳市	151
一、支撑背景:大数据成为贵州全省战略行动	151
二、制度保障:领导重视与体制机制建设	152
三、法治基础:推动数据开放地方立法	153
四、上下联动:市区两级政府数据开放一体化	154
五、引领全国:地市级数据开放平台中的标杆	155
第三节 虽小先飞,先统筹后开放:佛山市南海区	157
一、背景条件:起步早,但烟囱林立	157
二、夯实基础:率先成立数据统筹局	158
三、厚积薄发:开放数据推动应用	160
第四节 行业先锋,政企合作开放交通数据:交通运输部	161
一、来龙去脉:“出行云”的由来	162
二、进展现状:“出行云”的功能	163
三、发挥效益:“出行云”产生的效果	165
四、发展历程:“出行云”是如何发展起来的	166
五、困境与挑战:开放意愿、数据治理与安全风险	168
六、未来路径:探索公共数据开放之路	170
第六章 我国政府开放数据为什么这么难?	171
第一节 机遇与压力:促成因素	171
一、国际标杆示范	171
二、地方自主探索	172
三、上级试点部署	172
四、同行学习比较	173
五、第三方评估	174
第二节 问题与纠结:阻碍因素	174
一、懂不懂开放?	174
二、愿不愿意开放?	175
三、可不可以开放?	176
四、能力够不够?	178
五、数据在哪里?	180

六、数据好不好? 180

七、平台建在哪一级? 182

八、谁在用开放数据? 183

第三节 一个盘根错节、不断生长的森林 183

第七章 中国政府数据开放路在何方? 186

第一节 扩大开放,枝杈相连 186

第二节 夯实基础,改善土壤 188

一、法规政策 188

二、领导支持 189

三、组织管理 189

四、人员能力 190

五、资金保障 190

六、数据治理 191

第三节 运行平台,提升体验 192

一、平台概览 192

二、平台导引 192

三、数据获取 192

四、工具提供 192

五、利用成果展示 193

六、互动交流 193

七、个性化整合 193

第四节 释放数据,确保质量 193

一、免费原则 193

二、数据数量 194

三、数据质量 194

四、数据标准 195

五、数据覆盖面 196

六、数据持续性 196

第五节 面向利用,动态循环 196

第六节 确保安全,分级分类 197

第七节 营造环境,培育生态 199

6 开放的数林:政府数据开放的中国故事

一、基础环境 199

二、利用能力 199

三、创新文化 200

第八节 整体路径与关键模块 200

第八章 一切才刚刚开始:政府数据开放的未来 203

第一节 不断变化的技术环境 203

一、物联网与数据开放 203

二、区块链与数据开放 204

三、人工智能与数据开放 205

第二节 变迁中的社会、经济与政治环境 206

一、法治伦理与数据开放 206

二、数据鸿沟与数据开放 207

三、全球治理与数据开放 208

第三节 组织变革与数据开放 209

第四节 开放数林的未来:大小两个生态系统的生长演化 210

索引 214

参考文献 224

后记 230

# 第一章 没有数据开放就没有智能时代

## ——政府数据开放的前世今生

2006年3月9日,英国《卫报》的文章《把皇冠上的明珠还给我们》指出:“我们纳税资助了政府部门去采集公共数据,但我们如果想要获得这些数据却需要再次付费,应该让这些数据免费开放,才能推动创新。”<sup>1</sup>这篇文章拉开了英国政府数据开放运动的序幕,也为世界政府数据开放的潮流奠定了基调。

之后,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八大基本原则:完整的、一手的、及时的、可获取的、可机读的、非歧视性的、非专属的、免授权的。到了2009年5月,美国联邦政府数据开放平台 Data.gov 上线运行,政府数据开放的潮流开始席卷全球。

然而,政府数据开放并非一朝一夕兴起的,而是在开放政府运动的引导下,在开源软件运动的助推下,并在大数据时代的呼唤下,才水到渠成、应运而生的。

### 第一节 政府数据开放的起源与背景

#### 一、封闭还是开放政府,这件事和数据有关

政府数据开放与开放政府运动密切相关,更确切地说,政府数据开放是开放政府运动发展到新阶段的产物。

政府应该是开放还是封闭,是人类千百年来争论不休的话题。有人强调政府与外部的界限以及高度受控的规则和秩序;有人则更强调价值多元以及个体自主决策的权利和能力。最终,公民在言论和出版等方面的权利被写进多数国家的宪法和法律中,成为开放政府理念的基石。<sup>2</sup>

20世纪五六十年代,一股政府改革的浪潮席卷世界,这其中就包括开放会议(Open Meeting)和信息自由(Freedom of Information)法案在各国的广泛推行。

这类法律又被称作“阳光法律”，旨在建立一种透明、参与和负责的政府文化，<sup>2</sup>其中较有代表性的法案包括1966年美国出台的《信息自由法案》。

近年来，随着政府数字化水平的不断提升，政府不再仅仅以纸质文件的方式生成和保存信息，而是越来越多地以电子化的形式采集和储存结构化的数据。“数据”是一手的原始记录，未经加工与解读，不具有明确意义，而“信息”是经过连接、加工或解读之后被赋予了意义的数字。<sup>3</sup>信息可供公众阅读，而数据不仅能被阅读，还能被开发利用。与此同时，公众对政府信息公开的要求也已不仅限于获知文件形式的信息，而是希望能获得政府数据并进行自由利用。传统方式的政府信息公开已无法满足公众需求，亟须拓展和深化，政府数据开放也由此登场。

前文提到的《卫报》文章就是关于英国陆地测量局的数据开放。文中提到，假如你为你的朋友买了一份报纸，然后你问这位朋友报纸上有些什么内容，这时他却要求你先付完钱才肯告诉你。这事听起来似乎很过分，但其实这件事的性质和政府部门在用纳税人的钱采集完数据后却不愿开放的做法是一样的。

## 二、从开放源代码到开放数据

“印刷文化”在欧洲的兴起曾极大促进了报纸、书籍等文化传播载体发展，最终带来了思想革命，<sup>2</sup>并影响到政府的组织结构和运作方式。从印刷术到互联网，信息通信技术的进步始终伴随着政府自身的转型发展，而近年来的开源软件运动也助推了政府数据开放。

1998年，受理查德·斯托曼(Richard Stallman)发起的“自由软件运动”<sup>4</sup>的影响，埃里克·雷蒙德(Eric Raymond)等人发起了“开源倡议”。<sup>5</sup>开放软件源代码作为一种新的软件开发方法，利用分布式同行评议和流程透明的力量，可使软件开发更优质，更可靠，更灵活和更省钱，并防止委托方被供应商剥削和套牢。<sup>6</sup>埃里克·雷蒙德在他的《大教堂与集市》一书中，把传统大型软件公司的开发模式形容为艰难而缓慢的大教堂建造过程，它采用了严密的管理方式和封闭的集中式结构，却在创新度、生产力和缺陷控制方面落后于开源软件的集市模式。与大教堂模式不同，集市模式是一种并行的、对等的、扁平化的开发结构，其参与者大多来自于互联网上的志愿者，结构松散，来去自由，看起来像是一个乱糟糟的集市，却取得了令人惊叹的成功。<sup>7</sup>开源的操作系统Linux和网页服务器Apache就是开源运动的成果。

得益于开源软件运动的成功，人们认识到了“开放”作为一种高效组织模式的

巨大潜力,从而为开放文化在其他领域的广泛扩散打下了基础。开放知识(Open Knowledge)、开放获取(Open Access)、开放硬件(Open Hardware)、开放货币(Open Currency)等各领域的开放运动此起彼伏。

这种开放文化也对政府治理方式产生了深刻影响,Web2.0 概念的首创者蒂姆·奥莱利(Tim O'Reilly)就认为,雷蒙德关于“大教堂与集市”的比喻也适用于政府治理。<sup>8</sup>例如,在传统的专利审查模式下,由于专利审查员人手不足和水平有限,导致专利申请积压和专利质量不断下降。因此,美国专利及商标局推出了公众专利评审项目(Peer to Patent),通过引入公众评审,释放了公众头脑中的“认知盈余”,使上述问题得到了缓解。纽约大学法学院教授贝丝·西蒙·诺维克(Beth Simone Noveck)由此系统地提出了“维基政府”的理念,<sup>9</sup>倡导通过应用信息技术和创新制度设计,来引导民众全程参与公共事务,吸纳集体智慧,形成群体智能,实现政府、公众、社会组织等多元主体的合作共治。

在开源运动的影响下,开放政府的理念也超越了信息公开和知情权的范畴,更加强调在开放数据基础上的参与和协作。政府开放数据给社会,供其进行增值开发和协作创新,正是政府数据开放的核心目标。开源运动的效用追求与开放政府的透明协作融汇并流,成为政府数据开放的重要源泉。

### 三、政府数据是智能时代的一种公共基础设施

进入大数据时代,人类获取、管理和利用数据的能力空前提升,社会各界对数据的价值愈发重视,希望政府开放数据的呼声也日益高涨。

大数据的发展一方面为政府数据的生成、开放和利用创造了更好的条件,另一方面,政府数据开放也将进一步推动大数据发展。大数据发展需要整合和利用各种来源的数据,而政府数据是其中不可或缺的重要组成部分。政府部门在履职过程中生成、获取和保存了大量基础性、关键性的数据资源,是一个国家最重要的数据保有者,在不涉及国家秘密、商业秘密、个人隐私的前提下,把政府数据开放给社会进行融合利用,将有力促进大数据和人工智能的发展。可以说,没有数据开放就没有大数据,没有大数据就没有智能时代,智能时代建立在开放数据的基础上。

在物理世界,政府在大型公共基础设施的建设中发挥了主导作用;在智能时代,数据作为一种新的基础设施,将和物理基础设施同等重要。开放政府数据,构筑公共数据基础设施,应是政府在大数据时代义不容辞的责任。