

大数据

政府统计的 新机遇

◎ 马建堂 主编

 中国统计出版社
China Statistics Press

大数据：政府统计的新机遇

马建堂 主编



中国统计出版社

China Statistics Press

图书在版编目(CIP)数据

大数据：政府统计的新机遇 / 马建堂主编. — 北京：中国统计出版社，2015.3

ISBN 978—7—5037—7483—6

I. ①大… II. ①马… III. ①国家行政机关—统计数据—数据处理—研究—中国 IV. ①C829.2

中国版本图书馆 CIP 数据核字(2015)第 162079 号

大数据：政府统计的新机遇

作者/马建堂

责任编辑/张 赏

装帧设计/张 冰

出版发行/中国统计出版社

通信地址/北京市丰台区西三环南路甲 6 号 邮政编码/100073

电 话/邮购(010)63376909 书店(010)68783171

网 址/<http://www.zgtjcbbs.com/>

印 刷/河北天普润印刷厂

经 销/新华书店

开 本/710×1000mm 1/16

字 数/180 千字

印 张/16

版 别/2015 年 8 月第 1 版

版 次/2015 年 8 月第 1 次印刷

定 价/49.00 元

版权所有。未经许可，本书的任何部分不得以任何方式在世界任何地区以任何文字翻印、拷贝、仿制或转载。

中国统计版图书，如有印装错误，本社发行部负责调换。

《大数据:政府统计的新机遇》

编委会

主 编 马建堂

副 主 编 鲜祖德

编写人员 万东华 孙学光

姜 澍 何 强 丛雅静 董 倩

施凤丹 耿蕴洁 宋 辉 谢 黎

朱国众 袁勤俭 庄燕杰

前 言

大数据正在催生一场史无前例的信息革命、产业革命和管理革命，深刻影响着经济社会生活和时代发展进程。伴随着新一代信息技术的快速发展，人类社会进入了一个数据量大、类型多样、变化加快、极具潜在价值的大数据时代。大数据正在对社会经济生活的各个层面产生深刻而持久的影响，催生了新的模式和新的业态，政府、企业和个人都在不同程度地享受着大数据带来的裨益。崭新的大数据时代，充满着机遇，也充满着挑战。

国家统计局适应大数据趋势，积极把握大数据机遇，将开发利用大数据作为重大战略，统筹谋划，系统研究，广泛合作，为政府统计应用大数据奠定了坚实基础。按照“总体设计、牵头攻关、先易后难、专业突破”的工作思路，积极推进大数据的探索与应用，努力打造统计数据来源“第二轨”。

大数据技术日新月异，大数据“宝库”的开发应用仍在路上。下阶段，政府统计将更加积极主动拥抱大数据，进一步加大大数据在政府统计中的开发利用力度。大数据背景下的政府统计，正在并将继续进行探索和完善。在数据获取环节，一方面，统计部门通过各种爬虫软件直接搜取各类大数据；另一方面，通过日臻成熟的数据采集技术和数据比对技术，从第三方渠道直接获取互联网上电商交易、行业平台、搜索引擎等多方面海量数据。同时，统计部门还将推动制定大数据应用的相关法律法规，积极促进各类主体大数据的开放共享，依法打造统计数据

来源的“第二轨”。在数据处理环节,将利用云计算,对搜集到的海量数据开展清洗和标准化处理,进行关联、验证、比对等分析,深度挖掘大数据价值。在数据应用环节,利用大数据评估校准传统结构化数据的同时,充分发挥大数据优势,生产出更及时、更全面、更准确、更直观的统计信息产品,更好地发挥服务型政府统计职能,提升政府统计工作效能,提高政府统计公信力。

为进一步深入研究、应用大数据,继 2013 年 10 月《大数据在政府统计中的探索与应用》出版之后,经过近两年的努力,我们又组织编写了《大数据:政府统计的新机遇》。本书主要介绍了国际组织、有关国家以及企业在大数据领域研究和应用的新进展,提出了大数据的来源、分类及标准,探讨了大数据的开放共享以及安全隐私保护,研究了大数据的可视化、空间大数据以及网络舆情大数据,阐述了国家统计局大数据发展战略,介绍了行政记录、商业记录、互联网大数据、物联网大数据、遥感大数据等在政府统计中的应用实践,并展望了大数据构建新型政府统计的美好前景。诚恳希望读者在了解大数据发展新态势、特别是在政府统计中应用新进展的基础上,反馈对本书的意见和建议。

我们正处于一个伟大的变革时代,谁拥有大数据,谁将占领最高点。大数据时代的来临,对政府统计既是挑战,更是机遇,大数据在政府统计中的探索与应用正在路上。中国政府统计,将努力适应新潮流,充分应用大数据,打造“第二轨”,使之与以结构化数据为主的传统之轨并行不悖、相辅相成,积极推动大数据相关标准和法律法规制定,大力促进数据互联互通、开放共享,不断提高统计能力、统计数据质量和政府统计公信力,不断提升中国政府统计的现代化水平!

马建堂

2015 年 3 月 18 日

目 录

前言:迎接大数据挑战

第1章 大数据新进展综述	1
第一节 国际组织大数据新动向	1
一、联合国	1
二、欧盟	2
三、世界经济论坛	3
四、开放政府联盟	4
第二节 国家大数据发展新战略	4
一、中国	4
二、美国	8
三、英国	9
四、日本	10
五、其他国家	11
第三节 大数据新进展述评	12
一、大数据引起大重视	12
二、大开放产生大价值	14
三、大数据推动大创新	15

四、深耕大数据技术要先行	15
五、利用大数据安全是前提	16
第2章 大数据来源、分类及标准	18
第一节 大数据的来源	18
一、传统数据源	19
二、新数据源	20
第二节 大数据的分类	23
一、行政记录数据	24
二、商业记录数据	26
三、互联网数据	28
第三节 大数据的标准	30
一、处在起步阶段的现状	30
二、创造规范统一的未来	31
第3章 大数据开放及实现	34
第一节 大数据开放的意义	34
一、数据开放是实现数据价值的基础	35
二、数据开放是推动社会创新的需要	37
第二节 国外经验及实际案例	39
一、发达国家(地区)的数据开放	39
二、社会组织、企业的数据开放	44
第三节 我国大数据开放及共享	46
一、我国数据开放的现状和问题	46
二、打造国家大数据开放平台	50

三、数据开放的保障机制	52
第4章 大数据安全与隐私保护	56
第一节 大数据时代面临的安全挑战	56
一、大数据用户隐私保护风险	56
二、大数据质量控制风险	59
三、大数据统计应用风险	61
第二节 大数据安全保障机制	63
一、确立数据归属权	63
二、构建数据应用规范	64
三、制定隐私保护法规	65
四、培育信息安全自律机制	67
五、加强网络安全宏观管理	68
第三节 大数据安全保护的技术手段	69
一、加密技术为大数据打造“金钟罩”	69
二、身份认证技术构建隐私保护新模式	69
三、数字水印技术提高大数据的可信度	71
四、数据溯源技术助力大数据应用	71
第5章 大数据可视化	73
第一节 大数据可视化概述	73
一、基本概念	73
二、可视化历史演进及发展	74
三、可视化的类型及基本原理	77
第二节 大数据可视化的操作流程	84

一、研究目的	85
二、数据采集	85
三、数据处理	86
四、数据挖掘	87
五、选择合适的可视化工具	89
六、可视化	90
七、美化、修饰与动画	90
第 6 章 空间大数据	93
第一节 空间大数据概述	94
一、空间不确定性	95
二、空间依赖性	96
三、空间异质性	97
第二节 空间大数据技术架构	99
一、空间大数据采集技术	100
二、空间大数据存储技术	101
三、空间大数据分析处理技术	102
四、空间大数据价值提取与展示技术	104
第三节 空间大数据应用前景	104
一、空间大数据网络建设	104
二、空间大数据计算挑战	105
三、空间大数据尺度问题	107
四、空间大数据隐私与安全	108
五、空间大数据能耗问题	110

第7章 网络舆情大数据	112
第一节 网络舆情和调查概述	112
一、基本概念	113
二、网络舆情的基本特点	114
三、网络舆情的类型与作用	115
四、网络舆情调查与方法	116
第二节 网络舆情大数据调查技术	118
一、大数据思维下的网络舆情调查	118
二、网络舆情大数据调查技术	119
三、网络舆情大数据调查基本步骤	121
第三节 网络舆情大数据调查的评价	122
一、网络舆情大数据调查的优势	122
二、网络舆情大数据调查应注意的问题	123
第8章 政府统计大数据发展战略	125
第一节 紧追大数据发展步伐加强顶层设计	126
第二节 倾力打造政府统计数据来源“第二轨”	128
第三节 先易后难 专业突破	131
第9章 行政记录在政府统计中的应用探索	133
第一节 行政记录概述	133
一、基本概念	133
二、行政记录分类	134
第二节 行政记录在政府统计的应用方式	135
一、直接应用	135

二、统计估算	135
三、抽样调查	136
四、统计加工	136
第三节 行政记录的政府统计应用实践	137
一、行政记录在基本单位统计中的应用	137
二、行政记录在常规统计中的应用	137
三、行政记录在普查中的应用	141
第四节 进一步加大行政记录大数据的应用力度	142
一、加强立法保证	142
二、完善统一编码系统和登记系统	143
三、加强行政机构的合作	144
第 10 章 商业记录在政府统计中的应用探索	145
第一节 商业记录概述	146
一、商业记录的来源和分类	146
二、商业记录对企业和政府统计的应用/价值差异	147
第二节 商业记录在政府统计的应用探索	148
一、电子商务商业记录在政府统计中的应用探索	149
二、企业生产经营管理商业记录在政府统计中的应用探索 ...	152
三、信息资讯商业记录在政府统计中的应用探索	156
第三节 商业记录在政府统计中应用展望	163
一、商业记录的应用途径及其统计标准分类	163
二、商业记录统计分析内容与模型方法	164

第 11 章 互联网大数据在政府统计中的应用探索	166
第一节 互联网大数据概述	166
第二节 互联网大数据对政府统计的应用价值	167
第三节 互联网大数据在政府统计中的应用实践	168
一、互联网大数据在预测房价方面的应用	169
二、互联网大数据在经济、人口等统计中的应用	173
第四节 互联网大数据应用展望	174
第 12 章 物联网大数据在政府统计中的应用探索	177
第一节 物联网及其数据特征概述	177
一、物联网的概念	177
二、物联网有别于传统互联网的特征	179
三、物联网大数据的特征	180
第二节 物联网在政府统计中的应用探索	182
一、物联网大数据在统计数据质量检验中的应用	182
二、射频识别技术在交通、产品和人口相关专业统计中的 应用	182
三、传感器网络对特定调查对象追踪和监控应用	184
四、物联网大数据在国民账户体系中的应用	185
五、物联网技术在构建智能统计监测模型中的应用	186
六、物联网技术在其他领域中的应用	187
第 13 章 遥感大数据在政府统计中的应用实践	189
第一节 遥感大数据概述	190
一、遥感大数据的来源	190

二、遥感大数据的分类	191
三、遥感大数据的特点	192
第二节 遥感大数据在政府统计中的应用价值	193
一、遥感大数据丰富了政府统计的数据源	195
二、遥感大数据有利于提高政府统计数据质量	196
三、遥感大数据有利于政府统计的可视化	197
四、遥感大数据有利于提高统计工作效率	198
第三节 遥感大数据的政府统计应用实践	198
一、投资项目遥感核查	199
二、农作物遥感测量	201
第四节 遥感大数据应用展望	207
一、传统调查数据和遥感大数据将共同存在	207
二、以“统计云”为依托整合政府部门遥感大数据	208
三、制定和完善遥感统计标准与规范体系	209
四、研究和推广遥感大数据关键技术	210
五、扩大“3S”技术在政府统计中的应用领域	210
 第 14 章 大数据构建新型政府统计	211
第一节 政府统计助推中国数据思维新发展	211
第二节 借力大数据打造政府统计新平台	215
第三节 开启政府统计应用大数据的美好明天	217
 参考文献	222
 后 记	236

图 目

图 1:循“数”管理三部曲	38
图 2:平行坐标系	79
图 3:Radviz 方法	79
图 4:散点图矩阵	79
图 5:Andrews 曲线	80
图 6:面向像素的技术	80
图 7:Chernoff 面法	81
图 8:星绘法	81
图 9:多线图	82
图 10:Survey Plot	82
图 11:维堆	83
图 12:树图	83
图 13:嵌套坐标系	84
图 14:空间大数据技术架构	99
图 15:2014 年 2 月—2015 年 2 月“1 号店指数”和 CPI 的环比 涨幅	150
图 16:基于网络爬虫技术的价格指数计算系统架构	151
图 17:2010 年 1 月—2015 年 3 月上海大宗商品价格指数(月度) 走势图	158

图 18:2014 年 2 月 28 日—2015 年 2 月 28 日泛亚稀有金属价格 指数走势	160
图 19:2015 年上半年北京市各区县地表动土工程分布	200
图 20:基于遥感大数据的 6 省冬小麦空间分布	204
图 21:内业数据处理流程	205
图 22:2013 年黑龙江秋粮空间分布情况	207

表 目

表 1:大数据的分类	24
表 2:数据挖掘工具概览	88
表 3:不同类型的可视化工具	90
表 4:2013 秋粮遥感测量 5 省份数据覆盖情况	206

专 栏

专栏 1:一天能产生多少数据	22
专栏 2:美国“棱镜门”事件	58
专栏 3:微故事:美国加州“橡皮擦法案”简介	64

第1章

大数据新进展综述

近几年，国际组织、各国政府以及知名专家学者，从不同角度对大数据进行了积极研究、开发与利用，推动着数据创新与社会进步。

第一节 国际组织大数据新动向

本节通过归纳总结有关国际组织在大数据领域的研究成果及主要观点，洞悉全球对大数据研究的新动态。

一、联合国

2012年5月，联合国“全球脉动”(Global Pulse)计划项目发布