

数 ■ 字 ■ 环 ■ 保 ■ 系 ■ 列 ■ 丛 ■ 书



互联网时代的 环境大数据

刘锐 刘俊 谢涛 孙世友 姚新◎主编

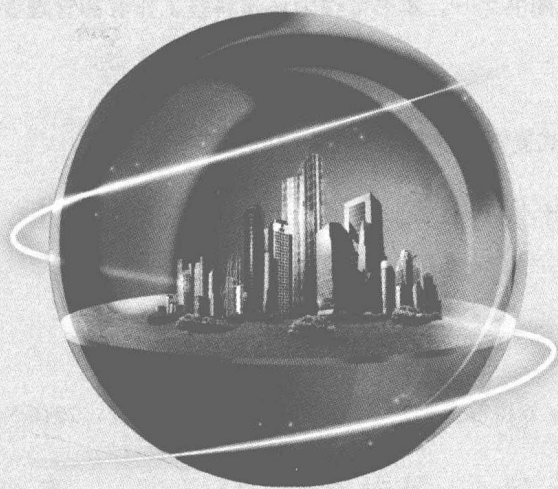


中国工信出版集团



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>

数·字·环·保·系·列·丛·书



互联网时代的 环境大数据

刘锐 刘俊 谢涛 孙世友 姚新 主编

电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

本书系统阐述了大数据与“互联网+”的内涵及发展历程、从环境大数据理论基础、环境大数据核心技术、环境大数据构成体系、环境大数据采集、环境大数据组织管理等方面对基于互联网的环境大数据进行总体论述。同时，阐述了环境大数据的应用、共享服务和保障机制等。另外，本书从实践角度介绍了我国环境大数据的应用案例和产业发展。

作为数字环保的系列图书之一，本书可作为环保科技工作者和管理者关于大数据技术在环保领域应用的参考著作。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有，侵权必究。

图书在版编目（CIP）数据

互联网时代的环境大数据 / 刘锐等主编. —北京：电子工业出版社，2016.5

ISBN 978-7-121-28434-2

I. ①互… II. ①刘… III. ①环境系统—数据处理IV. ①X192

中国版本图书馆 CIP 数据核字（2016）第 058445 号

策划编辑：吴长莘

责任编辑：郝黎明

印 刷：北京季蜂印刷有限公司

装 订：北京季蜂印刷有限公司

出版发行：电子工业出版社

北京市海淀区万寿路 173 信箱 邮编 100036

开 本：720×1 000 1/16 印张：27.25 字数：610.4 千字

版 次：2016 年 5 月第 1 版

印 次：2016 年 5 月第 1 次印刷

定 价：88.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888，88258888。

质量投诉请发邮件至 zltz@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式：（010）88254756。

编辑委员会

主 编

刘 锐 刘 俊 谢 涛 孙世友 姚 新

副主编

郭站君 房 明 顾伟伟 童 元 汪玉峰
李淼泉 张 林 刘艳民 侯立涛 卫晋晋

编 委

陈 静 李红华 许青云 刘耕源 沈小华
曹世凯 胡 珺 陈亚新 盛 琳 陈贝贝
王佳怡 郭 晓 张芬芬 徐春萌 赵杰迪
张 平 李 慧 樊俊佳 李志鹏 谭 淼
毛奎重 孔祥军 路淑斌 张慧良 杨 龙
王 健 徐 跃 马麒钧 邹 芬 王思怡
李月彬 唐小坤 陈泓熹 孙 阳

序 | PROLOGUE

我国政府一直高度重视环境质量的改善，党的“十七大”把环境保护摆上了重要的战略位置，党的“十八大”则把生态文明建设提升到“五位一体”总体布局的战略高度。“十八大”提出大力推进生态文明建设，建设美丽中国，实现中华民族永续发展。在互联网和移动互联网广泛应用的年代，环境质量数据量、污染源数据量和生态环境数据量迅猛增长，如何建设环境大数据，如何利用环境大数据为环境保护事业提供服务，是我阅读该书时最希望了解的答案。“互联网+环境保护”为新时代“智慧环保”体系建设提供有益参考，对于提高环境保护与管理决策的应用服务能力，推动互联网与生态文明建设深度融合具有十分重要的意义。

作为环境保护领域的新名词，许多学者对“互联网+环境保护”、环境大数据概念的理解仍比较抽象。该书系统阐述了“互联网+环境保护”与大数据的发展历程、理论基础、技术方法与应用案例等方面的知识，深入浅出地阐述了环境大数据的丰富内涵，并从实践角度介绍了我国环境大数据的产业发展。

“互联网+环境保护”是基于相关技术发展而提出的一系列技术的组合，如网络爬虫、分布式文件系统、数据挖掘、机器学习、环境模型、大数据可视化等，这些技术的发展都为环境大数据的建设提供了技术保障和支持，然而这些技术与环境大数据的结合还存在于理论与实践之间，离真正的落地实践还存在一定的距离。目前，我国在环境大数据发展和应用方面已具备一定的基础。然而，现阶段我国的环境大数据应用存在政府数据开放共享不足、缺乏顶层设计和统筹规划、法律法规建设滞后、创新应用领域不广等问题，各项工作基础相对薄弱。本书的出版，为指导我国在环境大数据建设与应用工作方面提供了有益参考。全书展现了“互联网+环境保护”建设的全貌，可为环保工作者的进一步探索开辟视野，是对环保工作经验的分享。该书内容丰富，技术科学、实用、可行，案例翔实，相信读者会从作者的论述中汲取到环境大数据建设丰富的理论方法和宝贵的先进经验。



2016年1月

致 谢 | CONVEY THANKS

环境大数据、“互联网+环境保护”是环境保护领域的新名词，其概念、理念都比较抽象，理论与实践之间还存在一定的距离。因此，本书参考了环境及大数据理论和案例的相关资料，涉及的单位包括环境保护部信息中心、河北省环境信息中心、中国电子技术标准化研究院、北京师范大学、微软（中国）有限公司、北京中科云谱物联网技术有限公司等，特此，对其单位和个人所提供的资料和帮助予以诚挚的感谢！

2016年1月

前 言 | PREFACE

信息技术与经济社会的交汇融合引发了数据迅猛增长。大数据是以容量大、类型多、存取速度快、应用价值高为主要特征的数据集合，正快速发展为对数量巨大、来源分散、格式多样的数据进行采集、存储和关联分析，从中发现新知识、创造新价值、提升新能力的新一代信息技术和服务业态。自 2009 年联合国制订了“数据脉动”计划开始，2010 年英国发起了“数据权”运动，2012 年美国实施了“大数据”战略，“大数据”时代的序幕渐渐拉开。而今，数据已成为国家基础性战略资源，大数据正日益对全球生产、流通、分配、消费活动，以及经济运行机制、社会生活方式和国家治理能力产生重要影响。全球范围内，运用大数据推动经济发展、完善社会治理、提升政府服务和监管能力正成为趋势，有关发达国家相继制定实施大数据战略性文件，大力推动大数据发展和应用。2014 年新加坡等地提出“大数据治国”理念；同年全国两会时，“大数据”第一次出现在政府工作报告中；2015 年 9 月，国务院印发《促进大数据发展行动纲要》，系统部署了我国的大数据发展工作。大数据已成为推动经济转型发展的新动力，重塑国家竞争优势的新机遇，也是提升政府治理能力的新途径。

目前，我国在大数据发展和应用方面已具备一定的基础。我国互联网、移动互联网用户规模居全球第一，拥有丰富的数据资源和应用市场优势，大数据部分关键技术研发取得突破，涌现出一批互联网创新企业和创新应用，一些地方政府已启动大数据相关工作。但同时也应该看到，现阶段我国的大数据应用存在政府数据开放共享不足、产业基础薄弱、缺乏顶层设计和统筹规划、法律法规建设滞后、创新应用领域不广等问题，亟待解决。坚持创新驱动发展，加快大数据部署，深化大数据应用，已成为稳增长、促改革、调结构、惠民生和推动政府治理能力现代化的内在需要和必然选择。

提到大数据，不得不提“互联网+”。“互联网+”是把互联网的创新成果与经济社会各领域深度融合，推动技术进步、效率提升和组织变革，提升实体经济创新力和生产力，形成更广泛的以互联网为基础设施和创新要素的经济社会发展新形态。在全球新一轮科技革命和产业变革中，互联网与各领域的融合发展具有广阔前景和无限潜力，已成为不可阻挡的时代潮流，正对各国经济社会发展产生着战略性和全局性的影响。国务院 2015 年 7 月 4 日印发了关于积极推进“互联网+”行动的指导意见。该意见明

确了我国推动“互联网+”的11项重点行动。其中，“互联网+”绿色生态重点行动中，明确要求“推动互联网与生态文明建设深度融合，完善污染物监测及信息发布系统，形成覆盖主要生态要素的资源环境承载能力动态监测网络，实现生态环境数据互联互通和开放共享。充分发挥互联网在逆向物流回收体系中的平台作用，促进再生资源交易利用便捷化、互动化、透明化，促进生产生活方式绿色化”。

在“互联网+”这一时代背景下，我们编写了《互联网时代的环境大数据》一书，旨在面向环境保护与管理决策的应用服务需要，探讨以大数据技术为驱动“互联网+环境保护”的技术体系与产业生态，为推动互联网与生态文明建设深度融合，建设新时代的“智慧环保”体系提供有益的参考。《互联网时代的环境大数据》作为数字环保的系列图书之一，是一本讲述环境大数据理论与实践的专业读物。本书共10章，分为4个部分。第一部分为第1~3章，为本书的绪论部分，重点阐述了互联网+、大数据、环境大数据的理念、内涵、基本概念，以及体系构成。第二部分为第4~7章，是本书的主体部分，重点阐述了环境大数据的数据采集、数据的组织管理、数据的分析应用和共享服务4个方面，这4个方面构成了环境大数据治理体系的核心。第三部分为第8章和第9章，重点阐述了环境大数据体系建设支撑与保障体系，设计安全、运行维护，以及部分体制、机制问题。第四部分为第10章，着重分析了“互联网+环境保护”的产业发展问题，探讨环境保护如何对打造大众创业、万众创新和增加公共产品、公共服务“双引擎”。

本书继承了中国科学院、中国环境科学研究院、北京师范大学、中科宇图科技股份有限公司、中科宇图资源环境科学研究院多年来在环境信息化领域的研究和实践成果。本书由中科宇图资源环境科学研究院、中科宇图科技股份有限公司、北京师范大学、中国科学院遥感与数字地球研究所等多家单位的专家共同编制。在此，对编写和出版本书做出贡献，以及关心本书的所有人员致以衷心的感谢！由于互联网+、环境大数据还是一个全新的领域，涉及的专业知识较广，加之编写时间仓促，书中难免有不足之处，敬请广大读者批评指正。

编者

2015年12月

目 录 | CONTENTS

第一章 “互联网+”与大数据	1
第一节 互联网时代与“互联网+”	2
一、互联网与互联网思维	2
二、“互联网+”的提出与时代要求	10
三、互联网思维与政府转型	25
第二节 “大数据”的定义、特征与应用	28
一、“大数据”与“大数据时代”	28
二、“大数据”的基本特征	29
三、“大数据”的发展形势和重要意义	31
四、现阶段我国大数据发展的目标与任务	33
第三节 大数据驱动“互联网+”	41
一、“互联网+”与物联网、云计算、大数据、工业4.0的关系	41
二、大数据对“互联网+”的驱动作用	44
第二章 环境大数据概述	47
第一节 环境大数据的相关概念	48
一、信息资源	48
二、环境信息	48
三、“大数据”与“环境大数据”	49
四、环境信息资源中心	50
第二节 环境大数据的来源	51
一、环境政务与业务数据	52
二、基于互联网和社会化获取的信息资源	57
第三节 发展环境大数据的作用与意义	58
第三章 环境大数据资源中心的构建及关键技术	60
第一节 环境大数据资源中心的构建	61
一、业务支撑	62
二、基础数据	64
三、基础设施	66
四、安全保障	67

第二节 环境大数据资源中心的关键技术	67
一、环境大数据基础平台的关键技术	69
二、环境大数据应用分析平台的关键技术	79
第四章 环境大数据的采集体系	94
第一节 基于物联网的环境数据多元感知体系	95
一、卫星遥感环境监测数据	95
二、航空遥感环境监测数据	106
三、地面环境监测数据	114
第二节 环境政务与业务数据集成	127
一、环境大数据集成需求	127
二、环境大数据集成目标	128
三、环境大数据集成类型	129
四、环境大数据集成思路	131
五、环境大数据集成方案	135
第三节 互联网环境数据信息的汇聚	141
一、互联网信息的环境数据来源	142
二、互联网环境信息的采集方案	142
第四节 环境大数据采集、传输及处理量估算	151
一、采集信息量存储估算	151
二、数据处理能力估算	158
三、应用系统数据存储估算	161
四、高性能计算集群运算能力估算	163
五、基础通信接入网络宽带估算	163
第五章 环境大数据的组织管理	165
第一节 环境大数据的数据标准	166
一、国内外大数据标准化工作进展	166
二、环境大数据标准化工作现状	171
三、环境大数据标准体系	172
四、未来环境大数据标准建设的重点工作	177
第二节 环境大数据资源目录体系构建	178
一、环境大数据资源目录体系概述	178
二、环境大数据资源目录体系分类	181
三、环境大数据资源目录体系支撑技术	186
四、环境大数据资源目录体系技术架构	188
第三节 环境大数据的管理	189

一、环境大数据管理的意义	189
二、环境大数据的存储	190
三、环境大数据的分析与应用	191
四、环境大数据的更新与维护	192
第六章 环境大数据的分析与应用	194
第一节 大数据在环保领域的应用需求	195
一、环境管理工作对环境大数据的需求	195
二、环境科技工作对环境大数据的需求	198
三、环境产业发展对环境大数据的需求	200
四、公众参与对环境大数据的需求	201
第二节 环境大数据在环境监测中的应用实例	202
一、环境监测面临的问题	202
二、监测系统架构和硬件	203
三、云端矫正技术框架	206
四、小结与展望	210
第三节 环境大数据在环境评价中的应用实例	210
一、水环境大数据分析背景概述	210
二、水环境大数据研究方法与技术路线	211
三、系统应用实例	212
四、结论	220
第四节 环境大数据在环境预测中的应用实例	220
一、环境预测的现状和问题	220
二、预测系统的架构	221
三、小结与展望	227
第五节 环境大数据在工业能源分析中的应用实例	229
一、工业能源环境大数据分析背景概述	229
二、大数据分析模型开发	230
三、基于大数据的工业能效评估模型的构建	235
四、基于大数据的 K-means 聚类分析的应用与实现	237
五、小结与展望	245
第七章 环境大数据的服务	247
第一节 环境大数据的服务概述	248
第二节 环境大数据的服务需求	250
一、国家战略的需求	250
二、环保行业部门的需求	254

三、行业企业的需求	257
四、科学研究的要求	260
五、社会数据的公众化需求	268
第三节 环境大数据的服务产品	274
一、一次数据产品	274
二、二次数据产品	275
三、环境保护指示性指标产品	277
四、环境大数据产品列表	278
第四节 环境大数据的服务方式	280
一、环境管理业务数据共享服务	280
二、环境专业领域百科信息服务	282
三、环境保护决策支持服务	285
第五节 环境大数据的商业模式	286
一、环境大数据的价值主张	286
二、环境大数据的客户细分	288
三、环境大数据的关键业务	290
四、环境大数据的核心资源	292
五、环境大数据的成本结构	294
第六节 环境大数据服务平台建设案例	296
一、国内外发展现状与趋势	296
二、环境大数据服务平台建设思路	298
三、环境大数据服务平台建设内容	299
第八章 环境大数据的网络硬件支撑与信息安全保障体系	304
第一节 环境大数据硬件支撑体系	305
一、物理资源建设	305
二、虚拟资源建设	311
三、整体管理架构	313
第二节 信息安全保障体系	319
一、加强信息安全建设的重要意义	319
二、信息安全管理体系统建设的具体目标	323
三、信息安全建设方法论	324
四、充分的现状调研和风险评估过程	326
五、设计建立信息安全保障体系总体框架	327
六、设计建立信息安全保障体系组织架构	328
七、设计建立信息安全保障体系管理体系	329

第九章 环境大数据资源中心业务化运行的保障机制	331
第一节 CIO 制度	334
一、CIO 概述	334
二、国内外 CIO 制度的发展现状	340
三、CIO 的组织结构	344
四、CIO 与环境大数据的发展变革	349
第二节 环境信息资源共享共建机制	350
一、环境数据共享与服务体系概述	350
二、必要性与意义	354
三、环境数据共享与服务体系技术框架和运营模式设计	357
第三节 环境大数据的质量保障与审核机制	359
一、环境大数据的质量保障机制	359
二、环境大数据的审核机制	370
第十章 环境大数据的产业发展	378
第一节 大数据产业发展概况	379
一、大数据产业概述	379
二、大数据产业链解析	380
三、全球大数据产业发展现状与预测	382
四、我国大数据产业发展现状与预测	388
五、我国大数据行业的误区	398
第二节 环境大数据产业发展现状	401
一、环境大数据应用“含苞待放”	401
二、大数据助力环境保护	406
三、大数据平台实现环境监测智能化	408
四、环保领域迎来大数据互联时代	409
第三节 环境大数据产业发展存在的问题	411
第四节 环境大数据产业发展展望	413
一、环境大数据可以解决的问题	413
二、开展环境保护大数据应用的基本思路	414
参考文献	417

第一章

“互联网+”与大数据

第一节 互联网时代与“互联网+”

一、互联网与互联网思维

(一) 互联网与移动互联网

1. 互联网

18 世纪蒸汽机的出现和广泛使用引发了第一次工业革命，用机器代替了手工劳动。19 世纪电力的大规模应用造就了第二次工业革命。电力革命给人类社会带来了巨大的进步。从 1986 年中国发出第一封国际电子邮件开始，经过近 30 年的发展，互联网在中国有了 6.49 亿用户，渗透率达到了 47.9%，互联网逐步从城市向农村渗透。现如今，互联网已经如电力般渗透到我们日常工作生活的每一个角落。

互联网受欢迎的根本原因在于它的成本低，其优点如下：①互联网能够不受空间限制来进行信息交换；②信息交换具有时域性（更新速度快）；③交换信息具有互动性（人与人、人与信息之间可以互动交流）；④信息交换的使用成本低（通过信息交换，代替实物交换）；⑤信息交换趋向于个性化发展（容易满足每个人的个性化需求）；⑥使用者众多；⑦有价值的信息被资源整合，信息储存量大、高效；⑧信息交换能以多种形式存在（视频、图片、文章等）。

那么什么是“互联网”呢？互联网（internet）又称网际网络，或音译为因特网（Internet）、英特网，是网络与网络之间所串连成的庞大网络，这些网络以一组通用的协议相连，形成逻辑上的单一巨大国际网络。通常 internet 泛指互联网，而 Internet 则特指因特网。这种将计算机网络互相连接在一起的方法可称作“网络互联”，在这基础上发展出覆盖全世界的全球性互联网络称互联网，即是互相连接在一起的网络结构。互联网并不等同于万维网，万维网只是一建基于超文本相互

链接而成的全球性系统，且是互联网所能提供服务的其中之一。

普遍认为互联网始于 1969 年的美国，是美军在 ARPA（阿帕网，美国国防部研究计划署）制定的协定下，首先用于军事连接，后将美国西南部的加利福尼亚大学洛杉矶分校、斯坦福大学研究学院、UCSB（加利福尼亚大学）和犹他州大学的 4 台主要的计算机连接起来。这个协定由剑桥大学的 BBN 和 MA 执行，在 1969 年 12 月开始联机。

从发展历程上看，互联网经历了如下几个阶段。

(1) 1961 年，美国麻省理工学院的 Leonard Kleinrock 博士发表了分组交换技术的论文，该技术后来演变为互联网的标准通信方式。

(2) 1968 年，参议员 Ted Kennedy 听说 BBN 赢得了 ARPA 协定作为内部消息处理器（IMP），Ted Kennedy 向 BBN 发送贺电祝贺他们在赢得“内部消息处理器”协议中表现出的精神。

(3) 1969 年，美国国防部启动具有抗核打击性计算机网络开发计划“ARPANET”。通信协议为 NCP（网络控制协议）。

(4) 1971 年，ARPANET 的技术开始向大学灯光科研机构普及。

(5) 1978 年，UUCP（UNIX 至 UNIX 复制协议）在贝尔实验室被提出来，1979 年，在 UUCP 的基础上新闻组网络系统发展起来。新闻组（集中某一主题的讨论组）紧跟着发展起来，它为在全世界范围内交换信息提供了一个新的方法。然而，新闻组并不被认为是互联网的一部分，因为它并不共享 TCP/IP 协议，它连接着遍布世界的 UNIX 系统，并且很多互联网站点都充分地利用新闻组。新闻组是网络世界发展中的非常重大的一部分。

(6) 1983 年，ARPANET 宣布旧通信协议 NCP（网络控制协议）

向 TCP/IP 过渡。正是因为采用具有扩展性的通信协议 TCP/IP，才能够将不同网络相互连接。因此，开发 TCP/IP 协议的 UCLA（加州大学洛杉矶分校）的学生 Vinton Cerf 等如今甚至被誉为“互联网之父”。

（7）1988 年，美国伊利诺斯大学的学生（当时）Steve Dorner 开始开发电子邮件软件“Eudora”。

（8）1989 年，第一个检索互联网的成就发明出来。它是由 Peter Deutsch 和他的全体成员在蒙特利尔的麦吉尔大学创造的，他们为 FTP 站点建立了一个档案，后来命名为 Archie。这个软件能周期性地到达所有开放的文件下载站点，列出他们的文件并且建立一个可以检索的软件索引。检索 Archie 命令是 UNIX 命令，所以只有利用 UNIX 知识才能充分利用它的性能。McFill 大学，第一个拥有 Archie 的大学，发现每天从美国到加拿大的通信中有一半的通信量访问 Archie。学校关心的是管理程序能否支持这么大的通信流量，因此只好关闭外部的访问。幸运的是当时有很多 Archie 可以利用。Brewster Kahle 当时在 Thinking Machines（智能计算机）发明了 WAIS（广域网信息服务），能够检索一个数据库下的所有文件和允许文件检索。根据复杂程度和性能情况不同有很多版本，但最简单的可以让网上的任何人可以利用。在它的高峰期，智能计算机公司维护着在全世界范围内能被 WAIS 检索的超过 600 个数据库的线索，包括所有的在新闻组里的常见问题文件和所有的正在开发中的用于网络标准的论文文档等。和 Archie 一样，它的接口并不是很直观，所以要想很好地利用它，也得花费很多的时间和精力。

（9）1989 年，在普及互联网应用的历史上又一个重大的事件发生了。Tim Berners 和其他在欧洲粒子物理实验室的人（这些人在欧洲粒子物理研究所非常出名）提出了一个分类互联网信息的协议。这个协议，1991 年后称为 WWW（World Wide Web），基于超文本协议——在一个文字中嵌入另一段文字的连接的系统，当你阅读这些页面的时候，