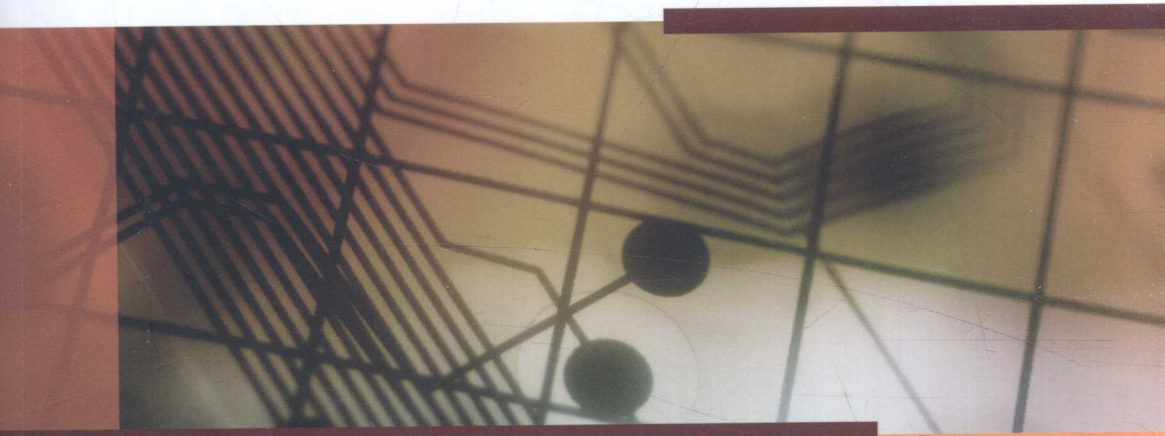
关联政府数据原理与应用

——大数据时代开放数据的技术与实践



Principle and Application of Linked Government Data

Technology and Practice of Open Data in Big Data Era

翟军 著

中国工信出版集团

電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

关联政府数据原理与应用

——大数据时代开放数据的技术与实践

翟　军　著

電子工業出版社
Publishing House of Electronics Industry
北京 · BEIJING

内 容 简 介

本书较为系统、深入地介绍了通过关联数据(Web 3.0)开放政府数据的主要流程和关键技术,包括数据的语义模型、元数据、URI 设计、数据查询和使用等。借助对 HTTP 重定向和内容协商的剖析,较为明晰地阐述了关联数据的基本原则。贯穿全书的开放数据实例和 Java 程序,能够很好地帮助读者理解所述内容。书中还包含开放数据的定义、开放数据网站、开放数据宪章、开放数据指数和五星模型等内容。

本书可供开放政府数据、关联开放数据和语义 Web 等领域研究和开发人员阅读和参考,也可作为相关专业(信息管理、电子政务、图书情报学、计算机应用等)高年级本科生和研究生的教材和参考书。对于政府开放数据的管理者和相关人员,也有一定的参考价值。

图书在版编目(CIP)数据

关联政府数据原理与应用:大数据时代开放数据的技术与实践/翟军著.
北京:电子工业出版社,2016.1
ISBN 978-7-121-27714-6

Ⅰ.①关… Ⅱ.①翟… Ⅲ.①电子政务-研究 Ⅳ.①D035.1-39

中国版本图书馆 CIP 数据核字(2015)第 287256 号

策划编辑:王志宇(wangzy@phei.com.cn)
责任编辑:王志宇
印　　刷:三河市双峰印刷装订有限公司
装　　订:三河市双峰印刷装订有限公司
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:16　字数:323 千字　插页:2
版　　次:2016 年 1 月第 1 版
印　　次:2016 年 1 月第 1 次印刷
定　　价:49.00 元

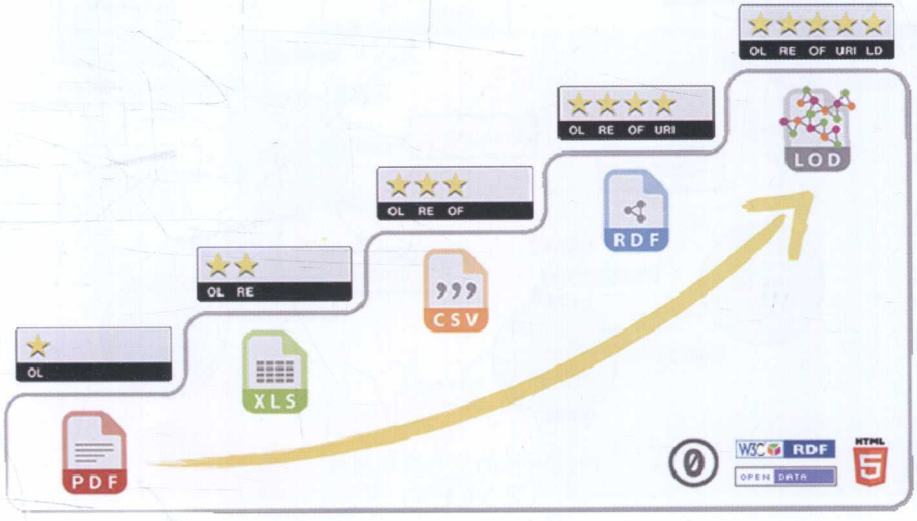
凡所购买电子工业出版社图书有缺损问题,请向购买书店调换。若书店售缺,请与本社发行部联系,联系及邮购电话:(010)88254888。

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

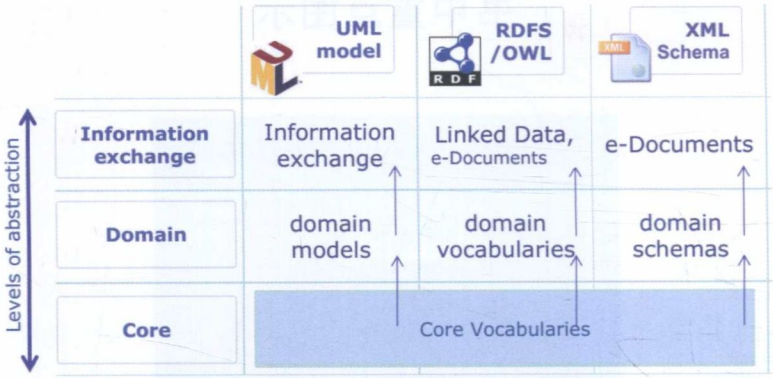
服务热线:(010)88258888。

书中重点图示

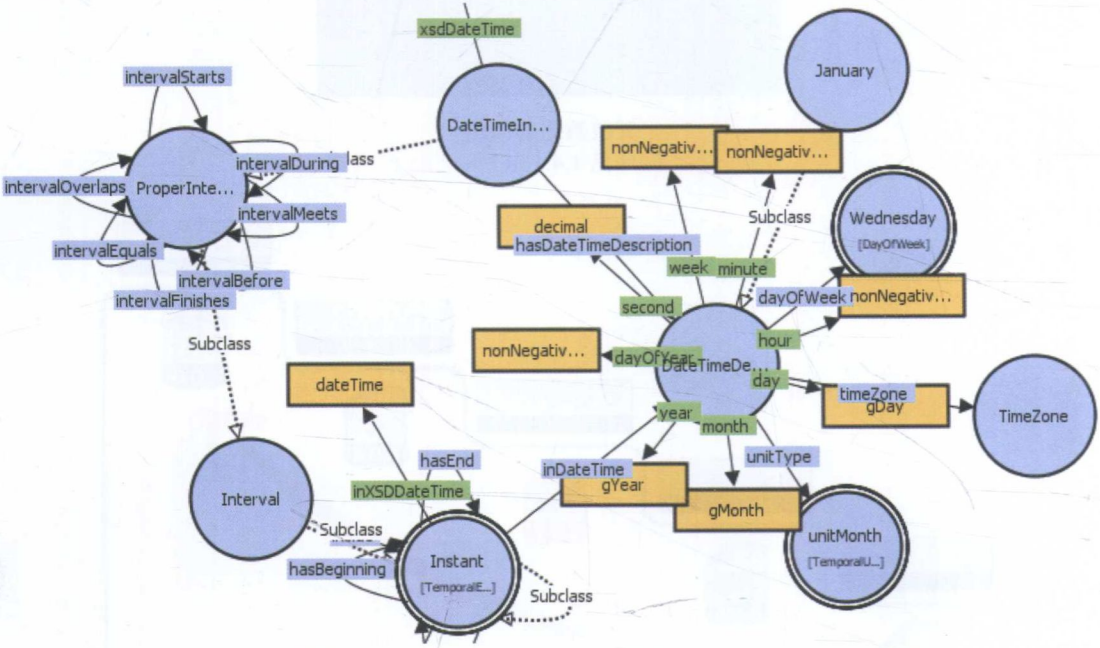
开放数据的冰山
（见 1.4.1 节）



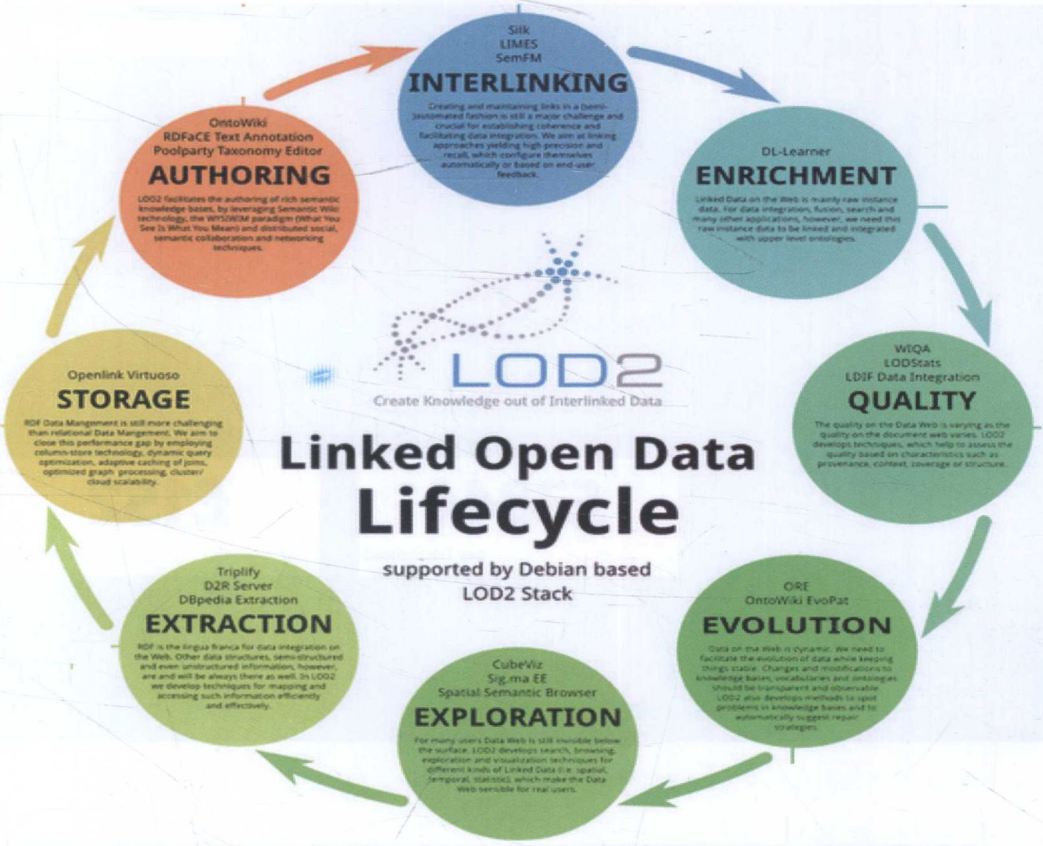
开放数据的五星评级模型
（见 1.5 节）



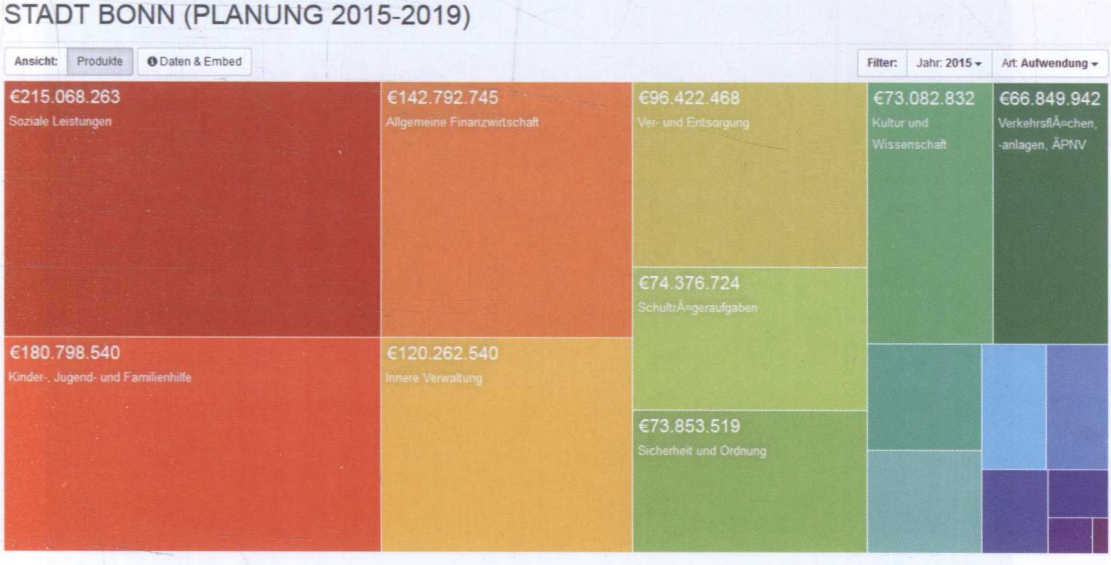
数据模型的形式化方法与抽象级别
（见 3.6.1 节）



时间本体的部分类和属性
（见 3.6.5 节）

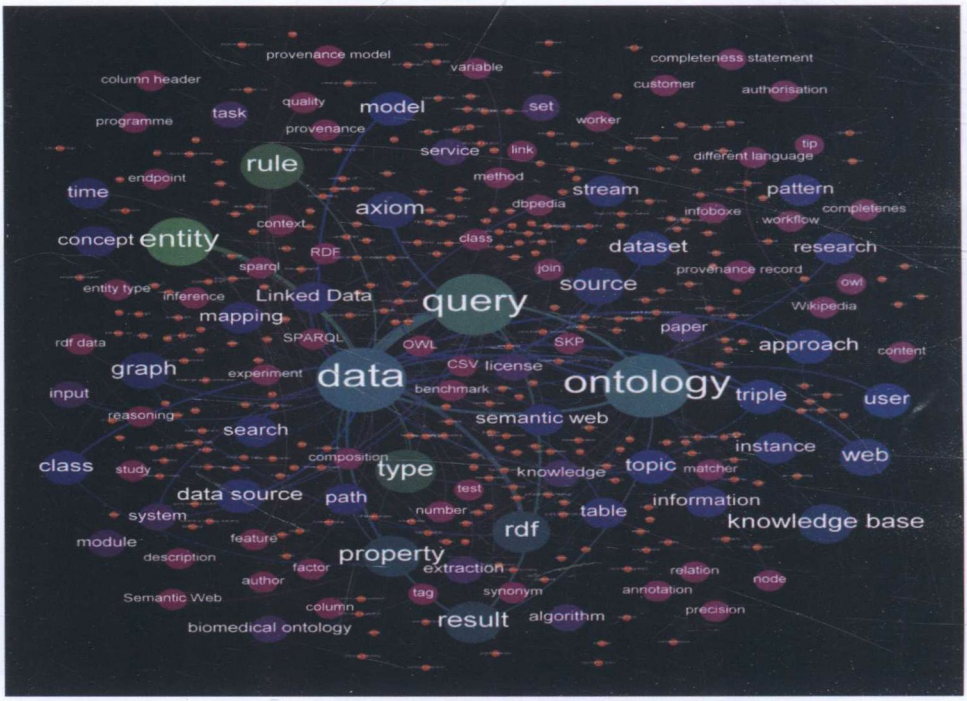


LOD2 项目支持的关联数据生命周期
（见 4.4 节）



波恩预算组成的可视化展示
（见 4.6.2 节）

南安普敦市的仪表盘
（见 9.4.2 节）



ISWC2013 发表论文的关键词网络
（见结束语）

前　言

2015年3月5日，李克强总理在《政府工作报告》中首次提出“互联网+”行动计划，涉及移动互联网、大数据、云计算、物联网等领域。其中，发展大数据产业是推动“互联网+”的必然需求。《新华（大连）软件和信息技术服务业发展指数报告（2015）》显示，大数据成为全球IT支出新增长点，2014年全球市场规模达到285亿美元。

大数据时代，“数据”的重要性被提到了前所未有的高度，通过对海量数据的交换、整合、分析和利用，能够发现新知识、创造新价值。在全社会中，政府数据起着核心与枢纽的作用。如果说数据资源是富有价值的“金矿”，英国著名的《卫报》则将公共数据（含政府数据）比喻为“皇冠上的明珠”。工程院院士潘云鹤在谈到“大数据是中国发展的一大机遇”时强调：“政府应在城市大数据的管理与开放中起主导作用。这主要表现在：促进知识服务业发展，创造新的市场与技术；确保个人信息不受侵犯、公共信息安全与共享；提高城市管理能力与决策水平，更好地为市民提供服务”。

各国的实践表明，开发和利用信息资源的前提是信息公开和数据开放，其核心是“开放政府数据”。我国专家也认为，数据的开放和跨界融合，是大数据产业得以发展壮大的关键。工程院院士、中国计算机学会大数据专家委员会主任李国杰认为：“数据的开放和共享是大数据时代国家治理体系现代化的前提”。智慧城市专家、工程院院士邬贺铨指出：“城市数据是智慧城市的重要资产，开放政府是智慧城市的前提，数据开放是评价开放政府的重要指标，开放数据将营造环境创新和释放商业机会”。全国政协委员、神州数码控股有限公司董事长郭为说：“只有共享的数据资源，才能释放数据的价值”。据估计，信息资源增值应用每年可为美国的医疗服务业节省3 000亿美元，为制造业在产品开发、组装等环节节省50%的成本，每年可为欧洲的公共部门管理节省2 500亿欧元，为全球个人位置数据服务提供商贡献1 000亿美元。麦肯锡全球研究所在2013年11月预测：开放数据在国际经济的一些领域，包括教育、交通、消费者产品、电力、石油、天然气、医疗保健、消费金融等行业的附加值逾3万亿美元，消费者也有希望获益于更大的价格透明度和获得更多的信息支持决策。2014年，澳大利亚咨询公司Lateral Economics的研究报告指出，综合G20各国的经济，开放数据将在未来5年实现总额为13万亿美元的增长，为G20国家贡献大约1.1个百分点的GDP增长，这将占到G20未来5年GDP增长目标（2%）的55%。

在此背景下，从2009年美国开始，越来越多的国家将政府数据开放作为国家战

略推动。到2015年，全球已有65个国家加入开放政府合作组织(OGP)，相继推出“国家行动计划”，从法律、技术、信息基础设施和应用等维度推动各层次的政府数据开放。

“开放数据”在欧盟的大数据发展战略中占有重要位置，包括资助“开放数据”领域的研究和创新活动、实施开放数据政策、促进公共数据的使用及再利用等。

我国也开展了信息公开和数据开放工作。中国政府网的数据栏目(www.gov.cn/shuju)、国家统计局的国家数据版块(data.stats.gov.cn)、环保部的数据中心(data-center.mep.gov.cn)都发布了各种统计数据。在大数据发展计划和智慧城市建设中，地方政府数据开放时代已经到来。2012—2015年间，北京、上海、贵州、浙江、武汉、青岛等在国内率先推出了数据开放门户网站。2012年2月，广东省宣布启动大数据战略，并在政府各部门开展数据开放试点，进一步推动政务公开。佛山市南海区打造的数据开放平台“数说·南海”，初步开放了48个单位的304个数据集，共14多万个数据记录。2015年8月—11月，上海市经济和信息化委员会、上海市交通委员会主办“上海开放数据创新应用大赛”(SODA)，通过“开放数据、众创协作”建立一套引导、选拔、扶持、推广优秀项目的完整机制。2015年9月—2016年1月，威海市政府主办以“开放数据，创业点亮威海”为主题的“2015威海市互联网+数据开放创业创新大赛”。

2015年6月24日，李克强总理主持召开国务院常务会议，部署推进“互联网+”行动，通过《“互联网+”行动指导意见》，确定了“搭建互联网+开放共享平台、加强公共服务、开展政务等公共数据开放利用试点”等相关支持措施。2015年7月1日，国务院办公厅公布《关于运用大数据加强对市场主体服务和监管的若干意见》，重点任务包括“探索建立政府信息资源目录”、“进一步加大政府信息公开和数据开放力度”等。2015年8月19日，国务院常务会议通过《关于促进大数据发展的行动纲要》，推动政府信息系统和公共数据互联共享，以及优先推动交通、医疗、就业、社保等民生领域政府数据向社会开放。2015年9月5日印发的《促进大数据发展行动纲要》提出的总体目标和主要任务包括：2017年年底前形成跨部门数据资源共享共用格局，到2018年中央政府层面实现数据统一共享交换平台的全覆盖；建立政府部门和事业单位等公共机构数据资源清单，制定实施政府数据开放共享标准，制订数据开放计划；2018年年底前，建成国家政府数据统一开放平台；2020年年底前，逐步实现信用、交通、医疗等民生保障服务相关领域的政府数据集向社会开放。

但总体而言，我国各级政府的数据开放还有很大的发展空间。

根据《2014年联合国电子政务调查报告》，中国的电子政务发展指数(EGDI)为0.545，在193个成员国中位列第70名。EGDI的世界平均值为0.471 2，25个国家(13%)的EGDI为“非常高”(大于0.75)，其平均值为0.836 8。2014年的报告首次

关注了“开放政府数据”，通过调查问卷对开放数据的进展情况进行了评估，中国位于评估得分高于66.6%的50个国家之一。

为追踪各国政府开放数据的状态，英国开放知识基金会(OKF)每年发布全球的“开放数据指数”。2014年中国政府数据的开放程度为37%，在97个国家和地区的排名从2013年的36名下滑至57名。排在第一的英国的开放程度为97%。中国的不足表现在：数据往往受版权保护而无开放授权、没有机器可读的格式而使应用程序很难直接获取数据，即“数据仅为公开、尚未达到开放的标准”。

万维网基金会(W3F)在“万维网指数”(Web Index)之后又推出了“开放数据晴雨表”，对全球数据开放情况进行评估和排序。2014年共评估了86个国家，中国以28.12的得分位于第46位，较2013年有了较大进步(排名61、得分11.82)。中国在准备程度、实施情况和影响力三个方面的得分分别为52、24和19，而排在第一的英国的得分为98、100和100。

在由独立学术组织WJP公布的“2015年全球开放政府指数”中，中国以0.43分位列全部102个国家和地区中的第87名。得分最高的前三个国家是瑞典、新西兰和挪威，同获0.81分。在政府数据公开程度、知情权、公民参与及投诉机制四个方面，中国的得分分别是0.52、0.53、0.21、0.46，排名分别是33、56、102和82。

在开放政府数据运动中，无论是《G8开放数据宪章》，还是美国总统的行政命令，以及欧盟的公共部门信息(PSI)再利用指令和我国的《大数据发展行动纲要》，都将Web作为数据开放的基础平台。在万维网发明25周年(1989—2014)之际，“欧洲信息学与数学研究联合会”会刊《ERCIM News》于2014年1月出版专刊“关联开放数据”，将其称为Web领域的“寂静的变革”。万维网之父蒂姆·伯纳斯-李认为下一代Web(Web 3.0)本质上是“关联数据万维网”(Linked Data Web, LDW)，是开放数据的理想平台。他建议以“关联数据”的形式发布政府数据，并提出“五星”模型，勾画出迁移到“关联数据”的路线图。

在蒂姆·伯纳斯-李、W3C和各国政府的推动下，英国、美国、欧盟等的“关联政府数据”已蔚为大观，涵盖教育、交通、统计、地理信息、图书馆与数字遗产等领域。这方面的最佳实践具有指导意义和借鉴价值，包括：政府数据建模、选择和创建词汇表/本体、URI设计、RDF转换、发布API和SPARQL端点等，这正是本书的关注内容。当然，由于时间和水平的限制，本书不可能涉猎“关联政府数据”理论研究、技术开发和实践应用的方方面面，而是立足于基本原理的解析，为探索者扫除一些必须跨越的障碍；通过剖析英国、美国、欧盟和中国的一些应用案例，展示电子政务在这一领域的现状和发展趋势；最后，书中的Java程序示例，对有志于开放数据创新应用者，会有所裨益。

2014年8月，联合国秘书长潘基文发起成立独立的专家顾问组IEAG，探讨以

“数据革命”促进可持续发展的相关问题，向全世界发出了数据革命的动员报告。2015年6月10日，阿里数据经济研究中心(ADEC)成立，发布的《云计算开启信息经济2.0》报告指出，传统的“计算机+软件”范式将向“云计算+数据”范式转型；另一份报告《从IT到DT》认为，DT(数据技术)的快速发展已经对商业体系的创新展示出巨大的变革潜力，而数据驱动的、全新的商业形态，也在呼唤着DT时代的治理创新与社会生活变革。美国白宫“智能信息披露”工作小组组长(2011—2012年)、纽约大学GovLab实验室资深顾问乔尔·古林在《开放数据：如何从无处不在的免费数据中发掘创意和商机》一书中认为：“开放数据是继互联网之后，又一轮改变全球商业模式的信息化浪潮”。

“关联政府数据”有望站在浪潮之巅。

阿里巴巴集团副总裁、《大数据》一书的作者涂子沛认为：“大数据之‘大’，将不仅仅意味着数据之多，还意味着，每个数据都能在互联网上获得生命、产生智能、散发活力和光彩”。为实现这一蓝图，关联数据和数据万维网(Web of Data)的作用是不可替代的。

谨以本书作为“关联开放数据”技术的入门之作。

本书得到了教育部科学技术研究重点项目(209030)、辽宁省教育厅项目(20060083、WT2010002)的部分资助；在写作过程中，参阅了国内外的大量文献、资料和素材，在此一并表示衷心感谢。同时，也感谢家人、同事和研究生对我写作和研究的支持和帮助。

书中难免存在错误、疏漏和不足，恳请各位读者不吝赐教。

作　者

目　录