

数据治理与数字化转型丛书

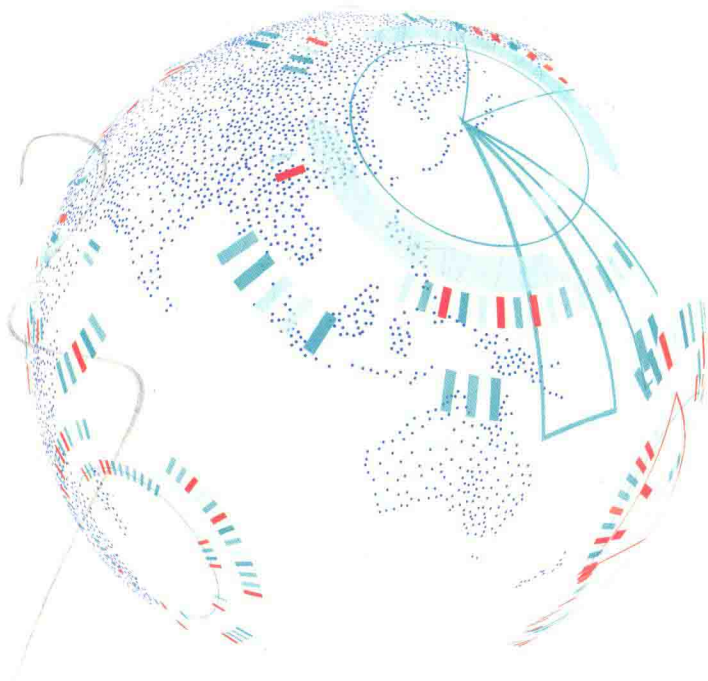
CCID 赛迪

DATA GOVERNANCE AND DATA SECURITY

数据治理与数据安全

张莉◎主编

中国电子信息产业发展研究院◎编著



中国工信出版集团



人民邮电出版社
POSTS & TELECOM PRESS

数据治理与数字化转型丛书



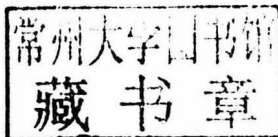
CCID赛迪

DATA GOVERNANCE
AND DATA SECURITY

数据治理与数据安全

张莉◎主编

中国电子信息产业发展研究院◎编著



人民邮电出版社
北京

图书在版编目 (C I P) 数据

数据治理与数据安全 / 张莉主编 ; 中国电子信息产业发展研究院编著. -- 北京 : 人民邮电出版社, 2019.9
ISBN 978-7-115-51590-2

I. ①数… II. ①张… ②中… III. ①数据处理—安全技术 IV. ①TP274

中国版本图书馆CIP数据核字(2019)第128189号

内 容 提 要

随着数字经济时代的到来,数据的价值不断被发掘。但是,围绕数据价值的挖掘也出现了诸多问题。例如,因为数据过于海量以及数据产权不明确导致用户不会用,因数据使用和保护不当引起隐私泄露等安全事件,等等。本书正是围绕这些问题,以“数据治理与数据安全”为主题展开论述。

首先,本书对数据治理的对象、主题、框架和方式等进行分析,使读者认识到数字经济时代数据流动的重要性和巨大意义。然后,本书从国内、国际两个方面进行阐述,力图为政府和企业开展个人信息保护、推进数据开放共享及跨境流动战略出谋划策。最后,本书基于我国的实际情况,借鉴国际经验,针对数据开放共享中存在的问题提出了具体落地的数据治理策略。总之,本书具有很高的理论意义和应用推广价值。

本书适合政府机构、科研机构、企事业单位中从事数据治理与安全工作的人员及高等院校相关专业的师生阅读。

◆ 主 编 张 莉

编 著 中国电子信息产业发展研究院

责任编辑 张国才

责任印制 彭志环

◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路 11 号

邮编 100164 电子邮件 315@ptpress.com.cn

网址 <http://www.ptpress.com.cn>

大厂聚鑫印刷有限责任公司印刷

◆ 开本: 700×1000 1/16

印张: 17.5

2019年9月第1版

字数: 200千字

2019年9月河北第1次印刷

定价: 69.00 元

读者服务热线: (010)81055656 印装质量热线: (010)81055316

反盗版热线: (010)81055315

广告经营许可证: 京东工商广登字 20170147 号



我们
一起
解决
问题

编 委 会

主 任 卢 山

副主任 黄子河

成 员 刘 权 吴志刚 张 莉 毕 颖

程 光 刘 蓓 洪延青 叶郁菲

王 超 张博卿 魏书音 周千荷



当今世界，信息技术创新日新月异，数字化、网络化、智能化深入发展，在推动经济社会发展、促进国家治理体系和治理能力现代化、满足人民日益增长的美好生活需要方面发挥着越来越重要的作用。党的十九大描绘了决胜全面建成小康社会、开启全面建设社会主义现代化国家新征程、实现中华民族伟大复兴的宏伟蓝图，对建设网络强国、数字中国、智慧社会做出了战略部署。2017年12月8日，习近平总书记在主持中共中央政治局第二次集体学习时指出，“大数据是信息化发展的新阶段”。2018年4月22日，习近平总书记在致首届数字中国建设峰会的贺信中强调：“加快数字中国建设，就是要适应我国发展新的历史方位，全面贯彻新发展理念，以信息化培育新动能，用新动能推动新发展，以新发展创造新辉煌。”2018年5月26日，习近平总书记在致2018中国国际大数据产业博览会贺信中重申“全面实施国家大数据战略，助力中国经济从高速增长转向高质量发展”。当前，信息化、大数据、数字经济等高频词语已成为世界各国推动经济社会可持续发展的着力点和竞争点。

在数字经济的发展历程中，数据起到了核心和关键作用。《经济学人》杂志曾将数据比喻为“21世纪的石油”，数据的重要性不言而喻。但是，数据毕竟具有诸多不同于石油的特征。例如，不仅不稀缺，反而可再生；不仅不排他，反而可以多方利用；不仅价值不长久，反而具有时效性，等等。因此，对数据价值的

挖掘，必须有别于对石油等传统资源的利用方式。

实际上，人们对数据价值的认识也经历了由浅入深、由简单趋向复杂的过程。总体来看，这个认知过程主要分为三个阶段：第一阶段是数据资源阶段，数据是记录、反映现实世界的一种资源；第二阶段是数据资产阶段，数据不仅是一种资源，还是一种资产，是个人或企业资产的重要组成部分，是创造财富的基础；第三阶段是数据资本阶段，数据的资源和资产的特性得到进一步发挥，与价值进行结合，通过交易等各种流动方式，最终变为资本。

不过，无论数据是资源、资产还是资本，其价值发挥在于汇聚、打通及利用。用一句话形容，就是数据“活”于流动之中。近些年，业界学界兴起了“数据治理”一词，并衍生出一系列模型和框架。其归根结底就是要实现数据的流动，避免数据成为一滩滩“死水”，一个个“孤岛”，在互联互通中最大程度地挖掘和释放数据的价值。

那么，数据如何流动？靠什么流动？推动数据开放共享是核心！当前，数据流动是通过数据开放、数据交换和数据交易等方式实现的。其中，数据开放主要指占据全社会数据资源约80%的政府数据的开放共享，数据交换和数据交易主要指政府与企业、企业与企业之间的数据开放共享。不过，与石油市场已经建立了一套权属和边界都很清晰的成熟交易规则不同，数据交易赖以开展的基础和前提——数据产权，在目前还是一个说不清、道不明的新鲜事物。而且，对于数据开放共享过程中的“大数据杀熟”“千人千价”等数据滥用、个人信息严重泄露等数据安全问题，我国目前也缺乏有效的应对措施。这些都成为当前数据治理领域的热点和难点，也是本书探讨的重点。

以上都是国内方面的情况。在国际方面，因为数据的跨境流动，也引发了人们对数据主权、数据本地化等问题的热烈讨论。例如，2018年10月科技部针对基因信息违法出境做出处罚等事件，都是业界关注的焦点，也是本书呼吁社

会各界加强重视的重要部分。

从内容上看，本书共分为7章。

第1章从数据治理的基本概念入手，探讨了数据与大数据的区别，数据体现出的资源、资产、资本的价值，阐述了我们对数据治理概念的理解，涉及数据治理的对象、主体、框架和方式，尤其是探讨了数字经济时代数据流动的重要性和巨大意义。我们认为，数据治理的核心就是推动数据自由、安全地流动，以便最大程度地挖掘和释放数据价值。要促使数据流动，国内层面主要就是推动数据的开放共享，实现数据“聚”“通”“用”。

第2章描述了当前数据开放共享存在的一系列问题。例如，因为不了解或太了解数据价值而产生的“不愿”开放共享的心理，因为数据安全问题频发而产生的“不敢”开放共享的心理，因为数据产权、数据定价等问题不明确而导致的“不会”开放共享的心理，这些原因都是数据开放共享的阻碍和掣肘因素。

第3章针对当前业界由于数据产权模糊不清所发生的争夺数据等矛盾和冲突事件提出了三个问题，即数据是谁的？谁在用数据？数据收益归谁？这三个问题正是本书分析数据产权的三个维度。

第4章指出了数字经济时代数据在给人类生产、生活带来巨大便利的同时也诱发了很多问题。例如，商家使用“千人千价”“动态定价”及“大数据杀熟”等方法，以钻法律空子、打擦边球的不正当方式赚取巨额利润；不法分子利用黑客技术盗取个人信息，造成个人信息泄露、用户画像被恶意利用等。这些数据滥用和数据安全问题将成为影响数据价值释放的“绊脚石”。

第5章主要阐述了国际层面的数据流动，即数据跨境流动。这一章梳理了数据主权的概念，分析了当前很多国家推行的数据本地化政策对数据保护的作用，以及对经济发展的延缓，同时列举了基因信息违法出境的案例，说明数据跨境流动是一把双刃剑，需要客观看待。

第6章主要描述了当前世界主要大国在推动数据流动、探索数据治理方面做出的努力，重点从数据开放共享、个人信息保护和数据跨境流动等方面进行论述。

第7章基于我国的实际情况，借鉴国际经验，主要针对前文提出的问题，分别提出治理策略。

数据治理是一个全新的话题，业界学界都未有定论，本书的出版是我院在这个领域开展研究的第一步，未来我们将继续深入研究，推出更多相关的成果。当然，本书写作过程中，由于时间仓促，加上作者水平有限，书中难免有纰漏，恳请广大读者批评指正。

第1章 流数不腐：数据“活”于流动	1
1.1 理解“数据治理”	4
1.1.1 数据	4
1.1.2 治理	9
1.1.3 数据治理	10
1.2 数据流动连接你我他	14
1.2.1 数据采集：让数据汇聚到一起	14
1.2.2 数据分析：机器学习和深度挖掘	16
1.2.3 数据关联：因果关系 or 相关关系	21
1.2.4 数据质量：“清洗”后的数据更可靠	24
1.2.5 数据反垄断：避免数据孤岛的新手段	25
1.3 数据流动避不开的几个话题	28
1.3.1 数据开放共享	28
1.3.2 数据产权	31
1.3.3 个人信息保护	33
1.3.4 数据跨境流动	36
第2章 数据开放共享：“三无”背后的重重顾虑	39
2.1 “数据开放共享”概念的发展	42
2.2 推进数据开放共享在路上	43
2.3 数据开放共享的主要方式	49
2.3.1 数据开放	49
2.3.2 数据交换	51
2.3.3 数据交易	53

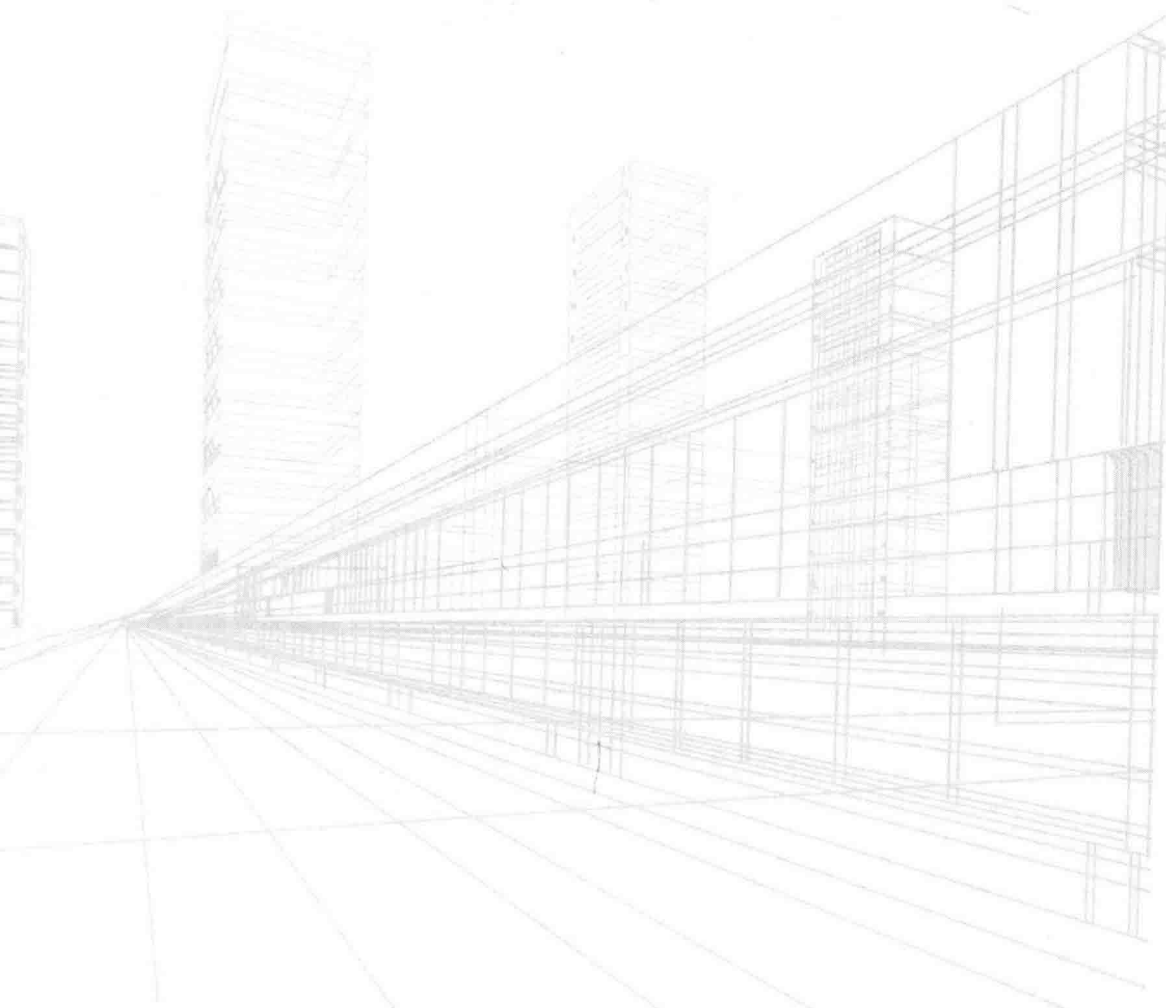
2.4	数据开放共享并非易事	55
2.4.1	无意愿开放共享	55
2.4.2	无胆量开放共享	58
2.4.3	无本领开放共享	62
第3章	数据产权：躲在被遗忘的角落里	65
3.1	数据产权：真命题还是伪命题	68
3.1.1	数据加工与数据资产化	68
3.1.2	用数据说话：关于数据争议案件的统计	71
3.1.3	数据产权的三大争议：所有权、使用权、收益权	74
3.2	数据产权面临的三大困境	78
3.2.1	从零开始的立法	78
3.2.2	捉襟见肘的司法实践	80
3.2.3	学术界的众说纷纭	83
第4章	数据滥用和安全事件频发：大数据发展的另一面	89
4.1	并非十全十美，算法也有局限性	92
4.1.1	认同、偏见与从众	92
4.1.2	算法局限：只让你看到认同的内容	93
4.1.3	算法歧视：公平性缺失愈发严重	96
4.2	价格操纵：数据成为“帮凶”	99
4.2.1	大数据杀熟	99
4.2.2	趁火打劫的动态定价	101
4.3	利益驱动：引发个人信息收集乱象的根源	105
4.3.1	用户画像：化成灰也认识你	105
4.3.2	默认勾选：“套取”个人信息的小技巧	108
4.3.3	霸王条款：“强取”个人信息的小策略	109
4.3.4	未经授权：“夺取”个人信息的小窍门	110
4.3.5	中小企业 App：容易被忽视的个人信息收集短板	111
4.4	数据安全：妨碍数据利用的最大变量	112
4.4.1	漏洞：数据泄露的重要源头	112
4.4.2	黑客：游走在漏洞边缘的逐利者	114

4.4.3	网络爬虫：数据泄露的新渠道	117
4.4.4	数据黑产：分工明确的数据利益链条	119
第5章	数据跨境流动：复杂与多变交织	125
5.1	数据主权与数据跨境流动	128
5.1.1	何为数据主权	128
5.1.2	云计算对数据主权的影响	132
5.1.3	数据跨境流动对数据主权的影响	134
5.2	数据本地化	137
5.2.1	数据本地化的几种形态	137
5.2.2	本地化能否实现数据安全	140
5.2.3	数据本地化的代价	141
5.3	数据跨境流动在我国	143
5.3.1	我国关于数据本地化存储的规定	143
5.3.2	由基因信息出境的案例说起	145
第6章	数据开放与保护：全球在行动	149
6.1	政府数据开放共享	152
6.1.1	联合国“全球脉动”项目	153
6.1.2	美国大数据战略	154
6.1.3	英国数据权运动	156
6.1.4	新加坡大数据治国理念	159
6.2	个人信息保护	161
6.2.1	全球数据保护立法聚焦个人信息保护	161
6.2.2	美国方案：从《隐私法案》到《加州消费者隐私法》	163
6.2.3	欧盟方案：从若干指令到《通用数据保护条例》	168
6.2.4	对我国的启示	175
6.3	数据跨境流动	178
6.3.1	美国方案：服务于贸易的宽松政策	178
6.3.2	欧盟方案：寻求个人信息保护与数据自由流动的 平衡	179
6.3.3	对我国的启示	188

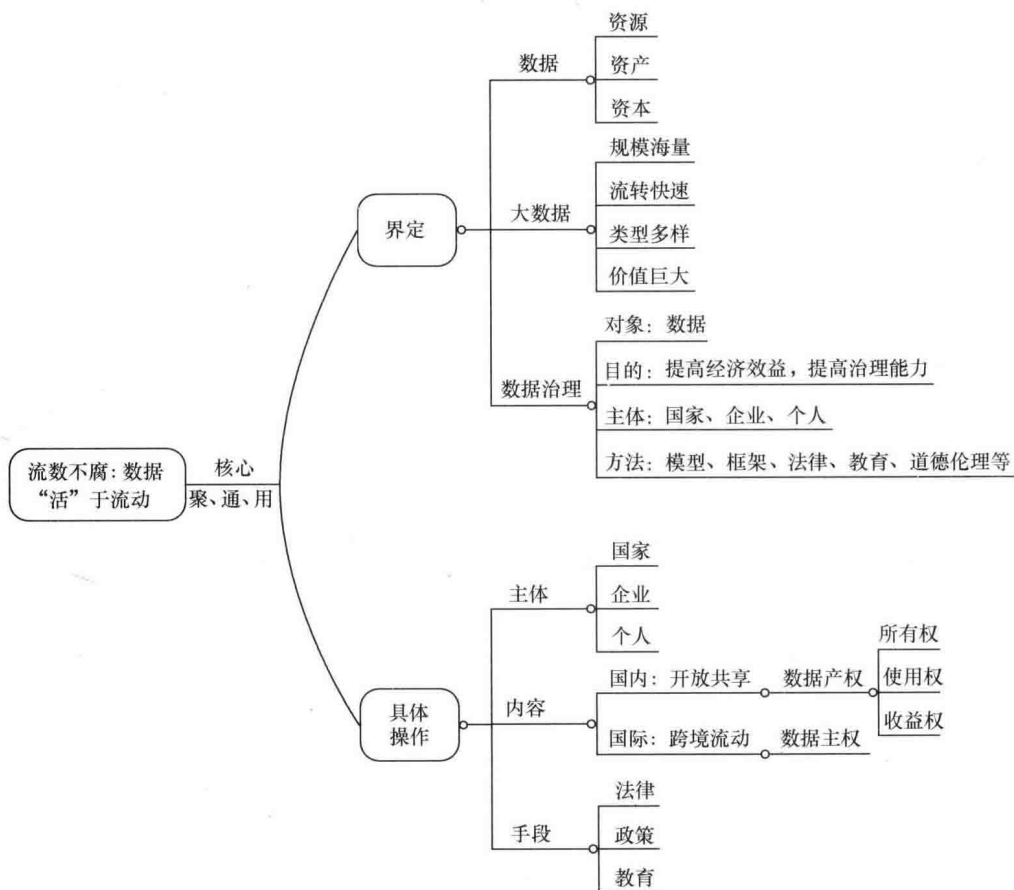
第7章 数据治理策略：基于四个维度	191
7.1 网民多、数据体量大	194
7.2 打通“数据孤岛”：基于数据开放共享的治理策略	196
7.2.1 提升各方对数据开放共享的认识	196
7.2.2 科学设计数据开放共享机制	198
7.2.3 构建安全框架，确保数据安全地开放共享	206
7.3 权能分离：基于数据产权的治理策略	211
7.3.1 数据产权的权利体系	212
7.3.2 数据产权的立法方向	215
7.3.3 数据所有权	217
7.3.4 数据使用权	222
7.3.5 数据收益权	231
7.4 保护个人信息：基于数据滥用和安全的治理策略	235
7.4.1 个人信息保护的现有举措	235
7.4.2 个人信息保护的痛点和难点	242
7.4.3 责任共担：国家、企业、个人	244
7.5 三个“不危害”：基于数据跨境流动的治理策略	254
7.5.1 我国数据跨境流动相关政策	254
7.5.2 数据跨境流动应坚持的原则	264
7.5.3 数据跨境流动，任重道远	266

第1章

流数不腐：
数据“活”于流动



思维导图



本章导读

本章主要聚焦“数据治理”的核心要义，从而引申出本书的全貌。文中梳理了当前业界学界对数据、治理以及数据治理的界定，分析了当前界定的不足之处，提出了我们关于数据治理核心要义的看法。我们认为，从一开始的数据资源到数据资产，再到当下流行的数据资本，人们对数据价值的认知是一个不断深化的过程，数据被赋予的内涵也更加丰富和复杂。与此同时，数据具有很多不同于石油等传统资源的特性。例如，传统资源是越用越少、不可再生，但数据是越积累量越大、越用越多。因此，对数据价值的挖掘，方式也应有别于传统做法。

我们认为，要最大程度地发挥数据的价值，根本在于促进其流动。无论各种主体以何种方式开展数据治理，其核心都是要推动数据自由、安全地流动，以便最大程度地挖掘和释放数据的价值。数据流动主要有两个层面，国内层面的抓手是推动数据开放共享，国际层面便是实现数据跨境流动。当前，阻碍数据顺畅流动的最主要掣肘因素有两个：一个是数据权问题，国内层面称为数据产权问题，国际层面称为数据主权问题；二是数据安全，目前国际社会主要聚焦于个人数据，即个人信息保护。如何看待和解决这些问题，给出何种治理策略，正是本书着力探讨的话题。

理解“数据治理”

随着大数据时代的到来，流动的数据已成为连接全世界的载体，也成为促进经济社会发展、便利人们生产生活的源动力。伴随着数据流动，尤其是为了解决流动过程中产生的一系列问题，“数据治理”一词逐渐兴起。而要了解数据治理，还得从数据、治理这些基本概念说起。

1.1.1 数据

数据与大数据

什么是数据？传统意义上，数据是指人类对事物进行测量的结果。如今，数据的概念有了很多延展。一般而言，数据是指对客观事件进行记录并可以鉴别的符号，是对客观事物的性质、状态以及相互关系等进行记载的物理符号或这些物理符号的组合。这些物理符号具有抽象、非随机的特点。

从数据的定义来看，数据具有两个特征，一个是差异性，另一个是规律性。差异性主要体现为数据多数描述的是事物的数量特征，现实世界中每件事、每个人、每种物都有不同的特征，因此反映于数据也会有各种不同的表现，甚至从表面看起来可能是杂乱无章的。规律性则主要体现为，数据是具有一定规律的，对数据进行分析研究，很重要的目的就是从数据中找出某种规律和关联。简而言之，正因为数据具有差异性，才有必要对数据进行研究与分析；也正因为数据存在规律性，对其研究才有价值。