

**THE DATA INDUSTRY:
THE BUSINESS
AND ECONOMICS
OF INFORMATION
AND BIG DATA**

CHUNLEI TANG

WILEY

CONTENTS

Preface

1 What is Data Industry?

- 1.1 Data, 2
 - 1.1.1 Data Resources, 3
 - 1.1.2 The Data Asset, 4
- 1.2 Industry, 6
 - 1.2.1 Industry Classification, 7
 - 1.2.2 The Modern Industrial System, 7
- 1.3 Data Industry, 10
 - 1.3.1 Definitions, 10
 - 1.3.2 An Industry Structure Study, 10
 - 1.3.3 Industrial Behavior, 13
 - 1.3.4 Market Performance, 16

2 Data Resources

- 2.1 Scientific Data, 19
 - 2.1.1 Data-Intensive Discovery in the Natural Sciences, 20
 - 2.1.2 The Social Sciences Revolution, 21
 - 2.1.3 The Underused Scientific Record, 22
- 2.2 Administrative Data, 22
 - 2.2.1 Open Governmental Affairs Data, 24
 - 2.2.2 Public Release of Administrative Data, 25
 - 2.2.3 A "Numerical" Misunderstanding in Governmental Affairs, 26

- 2.3 Internet Data, 26
 - 2.3.1 Cyberspace: Data of the Sole Existence, 27
 - 2.3.2 Crawled Fortune, 28
 - 2.3.3 Forum Opinion Mining, 28
 - 2.3.4 Chat with Hidden Identities, 29
 - 2.3.5 Email: The First Type of Electronic Evidence, 30
 - 2.3.6 Evolution of the Blog, 31
 - 2.3.7 Six Degrees Social Network, 32
- 2.4 Financial Data, 33
 - 2.4.1 Twins on News and Financial Data, 33
 - 2.4.2 The Annoyed Data Center, 33
- 2.5 Health Data, 34
 - 2.5.1 Clinical Data: EMRs, EHRs, and PHRs, 34
 - 2.5.2 Medicare Claims Data Fraud and Abuse Detection, 35
- 2.6 Transportation Data, 36
 - 2.6.1 Trajectory Data, 37
 - 2.6.2 Fixed-Position Data, 37
 - 2.6.3 Location-Based Data, 38
- 2.7 Transaction Data, 38
 - 2.7.1 Receipts Data, 39
 - 2.7.2 e-Commerce Data, 39

Data Industry Chain

- 3.1 Industrial Chain Definition, 41
 - 3.1.1 The Meaning and Characteristics, 41
 - 3.1.2 Attribute-Based Categories, 43
- 3.2 Industrial Chain Structure, 43
 - 3.2.1 Economic Entities, 44
 - 3.2.2 Environmental Elements, 44
- 3.3 Industrial Chain Formation, 46
 - 3.3.1 Value Analysis, 46
 - 3.3.2 Dimensional Matching, 50
- 3.4 Evolution of Industrial Chain, 51
- 3.5 Industrial Chain Governance, 53
 - 3.5.1 Governance Patterns, 53
 - 3.5.2 Instruments of Governance, 54
- 3.6 The Data Industry Chain and its Innovation Network, 56
 - 3.6.1 Innovation Layers, 56
 - 3.6.2 A Support System, 57

Existing Data Innovations

- 4.1 Web Creations, 59
 - 4.1.1 Network Writing, 60
 - 4.1.2 Creative Designs, 61

- 4.1.3 Bespoke Style, 62
- 4.1.4 Crowdsourcing, 63
- 4.2 Data Marketing, 63
 - 4.2.1 Market Positioning, 64
 - 4.2.2 Business Insights, 64
 - 4.2.3 Customer Evaluation, 66
- 4.3 Push Services, 67
 - 4.3.1 Targeted Advertising, 67
 - 4.3.2 Instant Broadcasting, 68
- 4.4 Price Comparison, 69
- 4.5 Disease Prevention, 70
 - 4.5.1 Tracking Epidemics, 71
 - 4.5.2 Whole-Genome Sequencing, 72

Data Services in Multiple Domains

- 5.1 Scientific Data Services, 73
 - 5.1.1 Literature Retrieval Reform, 74
 - 5.1.2 An Alternative Scholarly Communication Initiative, 74
 - 5.1.3 Scientific Research Project Services, 75
- 5.2 Administrative Data Services, 76
 - 5.2.1 Police Department, 77
 - 5.2.2 Statistical Office, 78
 - 5.2.3 Environmental Protection Agency, 78
- 5.3 Internet Data Services, 79
 - 5.3.1 Open Source, 79
 - 5.3.2 Privacy Services, 80
 - 5.3.3 People Search, 82
- 5.4 Financial Data Services, 82
 - 5.4.1 Describing Correlations, 83
 - 5.4.2 Simulating Market-Makers' Behaviors, 84
 - 5.4.3 Forecasting Security Prices, 85
- 5.5 Health Data Services, 86
 - 5.5.1 Approaching the Healthcare Singularity, 87
 - 5.5.2 New Drug of Launching Shortcuts, 87
 - 5.5.3 Monitoring in Chronic Disease, 88
 - 5.5.4 Data Supporting Data: Brain Sciences and Traditional Chinese Medicine, 90
- 5.6 Transportation Data Services, 91
 - 5.6.1 Household Travel Characteristics, 91
 - 5.6.2 Multivariate Analysis of Traffic Congestion, 92
 - 5.6.3 Short-Term Travel Time Estimation, 93
- 5.7 Transaction Data Services, 94
 - 5.7.1 Pricing Reform, 94
 - 5.7.2 Sales Transformation, 95
 - 5.7.3 Payment Upgrading, 96

6 Data Services in Distinct Sectors**99**

- 6.1 Natural Resource Sectors, 99
 - 6.1.1 Agriculture: Rely on What?, 100
 - 6.1.2 Forestry Sector: Grain for Green at All Costs?, 101
 - 6.1.3 Livestock and Poultry Sector: Making Early Warning to Be More Effective, 101
 - 6.1.4 Marine Sector: How to Support the Ocean Economy?, 102
 - 6.1.5 Extraction Sector: A New Exploration Strategy, 103
- 6.2 Manufacturing Sector, 104
 - 6.2.1 Production Capacity Optimization, 104
 - 6.2.2 Transforming the Production Process, 105
- 6.3 Logistics and Warehousing Sector, 106
 - 6.3.1 Optimizing Order Picking, 106
 - 6.3.2 Dynamic Equilibrium Logistic Channels, 107
- 6.4 Shipping Sector, 107
 - 6.4.1 Extracting More Transportation Capacity, 108
 - 6.4.2 Determining the Optimal Transfer in Road, Rail, Air, and Water Transport, 108
- 6.5 Real Estate Sector, 109
 - 6.5.1 Urban Planning: Along the Timeline, 109
 - 6.5.2 Commercial Layout: To Be Unique, 110
 - 6.5.3 Property Management: Become Intelligent, 110
- 6.6 Tourism Sector, 111
 - 6.6.1 Travel Arrangements, 111
 - 6.6.2 Pushing Attractions, 112
 - 6.6.3 Gourmet Food Recommendations, 112
 - 6.6.4 Accommodation Bidding, 113
- 6.7 Education and Training Sector, 113
 - 6.7.1 New Knowledge Appraisal Mechanism, 114
 - 6.7.2 Innovative Continuing Education, 114
- 6.8 Service Sector, 115
 - 6.8.1 Prolong Life: More Scientific, 115
 - 6.8.2 Elderly Care: Technology-Enhanced, Enough?, 116
 - 6.8.3 Legal Services: Occupational Changes, 117
 - 6.8.4 Patents: The Maximum Open Data Resource, 117
 - 6.8.5 Meteorological Data Services: How to Commercialize?, 118
- 6.9 Media, Sports, and the Entertainment Sector, 119
 - 6.9.1 Data Talent Scout, 119
 - 6.9.2 Interactive Script, 120
- 6.10 Public Sector, 121
 - 6.10.1 Wargaming, 121
 - 6.10.2 Public Opinion Analysis, 122

7	Business Models in the Data Industry	123
7.1	General Analysis of the Business Model, 123	
7.1.1	A Set of Elements and Their Relationships, 124	
7.1.2	Forming a Specific Business Logic, 125	
7.1.3	Creating and Commercializing Value, 125	
7.2	Data Industry Business Models, 126	
7.2.1	A Resource-Based View: Resource Possession, 126	
7.2.2	A Dynamic-Capability View: Endogenous Capacity, 127	
7.2.3	A Capital-Based View: Venture-Capital Operation, 128	
7.3	Innovation of Data Industry Business Models, 129	
7.3.1	Sources, 129	
7.3.2	Methods, 131	
7.3.3	A Paradox, 132	
8	Operating Models in the Data Industry	135
8.1	General Analysis of the Operating Model, 136	
8.1.1	Strategic Management, 136	
8.1.2	Competitiveness, 137	
8.1.3	Convergence, 137	
8.2	Data Industry Operating Models, 138	
8.2.1	Gradual Development: Google, 138	
8.2.2	Micro-Innovation: Baidu, 139	
8.2.3	Outsourcing: EMC, 140	
8.2.4	Data-Driven Restructuring: IBM, 140	
8.2.5	Mergers and Acquisitions: Yahoo!, 141	
8.2.6	Reengineering: Facebook, 142	
8.2.7	The Second Venture: Alibaba, 143	
8.3	Innovation of Data Industry Operating Models, 144	
8.3.1	Philosophy of Business, 144	
8.3.2	Management Styles, 145	
8.3.3	Force Field Analysis, 145	
9	Enterprise Agglomeration of the Data Industry	147
9.1	Directive Agglomeration, 148	
9.1.1	Data Resource Endowment, 148	
9.1.2	Multiple Target Sites, 149	
9.2	Driven Agglomeration, 149	
9.2.1	Labor Force, 150	
9.2.2	Capital, 150	
9.2.3	Technology, 151	
9.3	Industrial Symbiosis, 152	
9.3.1	Entity Symbiosis, 152	
9.3.2	Virtual Derivative, 153	

- 9.4 Wheel-Axle Type Agglomeration, 154
 - 9.4.1 Vertical Leadership Development, 154
 - 9.4.2 The Radiation Effect of Growth Poles, 154
- 9.5 Refocusing Agglomeration, 155
 - 9.5.1 "Smart Heart" of the Central Business District, 155
 - 9.5.2 The Core Objective "Besiege", 156

10 Cluster Effects of the Data Industry

159

- 10.1 External Economies, 159
 - 10.1.1 External Economies of Scale, 160
 - 10.1.2 External Economies of Scope, 160
- 10.2 Internal Economies, 161
 - 10.2.1 Coopetition, 161
 - 10.2.2 Synergy, 163
- 10.3 Transaction Cost, 164
 - 10.3.1 The Division of Cost, 164
 - 10.3.2 Opportunity Cost, 165
 - 10.3.3 Monitoring Cost, 166
- 10.4 Competitive Advantages, 167
 - 10.4.1 Innovation Performance, 167
 - 10.4.2 The Impact of Expansion, 168
- 10.5 Negative Effects, 169
 - 10.5.1 Innovation Risk, 169
 - 10.5.2 Data Asset Specificity, 169
 - 10.5.3 Crowding Effect, 170

11 A Mode of Industrial Development for the Data Industry

171

- 11.1 General Analysis of the Development Mode, 171
 - 11.1.1 Influence Factors, 172
 - 11.1.2 Dominant Styles, 172
- 11.2 A Basic Development Mode for the Data Industry, 173
 - 11.2.1 Industrial Structure: A Comprehensive Advancement Plan, 173
 - 11.2.2 Industrial Organization: Dominated by the SMEs, 174
 - 11.2.3 Industrial Distribution: Endogenous Growth, 174
 - 11.2.4 Industrial Strategy: Self-Dependent Innovation, 175
 - 11.2.5 Industrial Policy: Market Driven, 176
- 11.3 An Optimized Development Mode for the Data Industry, 176
 - 11.3.1 New Industrial Structure: Built on Upgrading of Traditional Industries, 176
 - 11.3.2 New Industrial Organization: Small Is Beautiful, 178
 - 11.3.3 New Industrial Distribution: Constructing a Novel Type of Industrial Bases, 178
 - 11.3.4 New Industrial Strategy: Industry/University Cooperation, 179
 - 11.3.5 New Industrial Policy: Civil-Military Coordination, 180

12 A Guide to the Emerging Data Law	183
12.1 Data Resource Law, 183	
12.2 Data Antitrust Law, 185	
12.3 Data Fraud Prevention Law, 186	
12.4 Data Privacy Law, 187	
12.5 Data Asset Law, 188	
References	189
Index	193