Contents

List of illustrations xii
Acknowledgements xiii
Introduction xiv

1 Legal data overview 1

1.1 Introduction 1
1.2 Why look at data? 2
1.3 Applications 3
Business improvement 4
Legal research 4
Academic research 5
Court processes 5
Consumer focused applications 6
1.4 Sources 6
Courts and tribunals 7
Legislative bodies 9
Government departments 11
Law firms and lawyers 11
Academic research 12
Legal technology companies 12
1.5 Global context 12
United States 12
Canada 13
United Kingdom 13
European Union 14
Africa 14
India 15
Singapore 15
Australia 16
1.6 Conclusion 16

2 Sources of data 19

2.1 Introduction 19
2.2 Sources 20
Governments 21
Parliamentary bodies 22
Courts 23
Business processes 26
Legal publishers and data providers 28
2.3 Developing data 29
Experimental methodology 29
Proxy data 30
2.4 Strategy 31
2.5 Conclusion 31

3 Data formats 34

3.1 Introduction 34
3.2 Categories 34
3.3 Issues particular to law 36
Case citations 36
Depth of data 36
Machine readable law 36
3.4 Types 37
Numerical 37
Categorical 38
Free text 38
3.5 Delivery formats 39
Spreadsheets 41
Tagged formats 41
Databases 42
APIs 42
3.6 Considerations 43
Data standards 43
Passively and actively collected data 43
Supervised and unsupervised machine learning 44
Law as code and law as data 45
3.7 Planning 46
3.8 Conclusion 47

4 Data analysis techniques 49

4.1 Introduction 49
4.2 Close reading 50

4.3 Statistical analysis 51
Normal distribution 52
Random distribution 52
Linear distribution 52
Multiple points of comparison 53
Techniques 53
Compared to machine learning 55
4.4 Machine learning 55
Supervised learning 56
Unsupervised learning 56
Reinforcement learning 57
4.5 Natural language processing 57
Bag-of-words 58
Vector analysis 59
4.6 Other tools 60
Knowledge bases 60
Decision trees 60
Network analysis 62
4.7 Conclusion 62

5 Interpreting legal data 65

5.1 Introduction 65
5.2 Approaching data analysis 66
5.3 Methodology 67
Statistics 67
Machine learning 69
Natural language 69
5.4 Considerations 70
Correlation and causation 70
Proxy data 71
Sensitivity 71
Suitability 72
Complexity 72
Extrapolation 73
5.5 Concerns 73
Overfitting 74
Bias 75
5.6 Conclusion 75

6 Issues with using legal data 78

6.1 Introduction 78

6.2 Availability 78
6.3 What is missing 79
6.4 Ambiguity 80
6.5 Limitations on language processing 81
6.6 Sampling 82
Analytics and sample size 82
6.7 Cost 83
6.8 Jurisdiction 84
6.9 Structure 84
Case law 85
Legislation 86
Law firms' data 87
6.10 Risks 88
6.11 Conclusion 89

7 Artificial intelligence 93

7.1 Introduction 93
7.2 Definition 93
7.3 Promise and limits 94
7.4 Types of artificial intelligence 95
Statistical analysis 95
Machine learning 96
Vaporware 97
7.5 Process 97
7.6 Likely impacts 98
7.7 Limitations 100
7.8 Concerns 100
Fairness 100
Error and bias 102
Ethics 102
7.9 Thinking critically 104
7.10 Conclusion 105

8 The law and politics of legal data 108

8.1 Introduction 108
8.2 The nature of data based applications 109
8.3 Access to data 110
8.4 Privacy 111
8.5 Jurisdiction 112

Infrastructure 112
Legacy systems 113
Government controls 114
8.6 Primary law 115
8.7 Practice data 116
8.8 The research landscape 116
8.9 Working toward change 117
8.10 Conclusion 118

9 Vision for the future 120

9.1 Introduction 120
9.2 Thinking about the future 121
9.3 Forecasting 122
9.4 Technology adoption dynamics 122
9.5 Issues 124
Data availability 124
Regulation of technology 126
Social concerns 127
9.6 Divergence 127
Jurisdiction 127
Population segments 128
Legal system 129
Culture 129
9.7 Predictions 130
Near term 130
Medium term 131
Long term 132
9.8 Conclusion 133

Bibliography 135
Index 146