

---

# Contents

---

<i>List of illustrations</i>	xii
<i>Acknowledgements</i>	xiii
<i>Introduction</i>	xiv

1	Legal data overview	1
1.1	<i>Introduction</i>	1
1.2	<i>Why look at data?</i>	2
1.3	<i>Applications</i>	3
	Business improvement	4
	Legal research	4
	Academic research	5
	Court processes	5
	Consumer focused applications	6
1.4	<i>Sources</i>	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	<i>Global context</i>	12
	United States	12
	Canada	13
	United Kingdom	13
	European Union	14
	Africa	14
	India	15
	Singapore	15
	Australia	16
1.6	<i>Conclusion</i>	16



2	Sources of data	19
2.1	<i>Introduction</i>	19
2.2	<i>Sources</i>	20
	Governments	21
	Parliamentary bodies	22
	Courts	23
	Business processes	26
	Legal publishers and data providers	28
2.3	<i>Developing data</i>	29
	Experimental methodology	29
	Proxy data	30
2.4	<i>Strategy</i>	31
2.5	<i>Conclusion</i>	31
3	Data formats	34
3.1	<i>Introduction</i>	34
3.2	<i>Categories</i>	34
3.3	<i>Issues particular to law</i>	36
	Case citations	36
	Depth of data	36
	Machine readable law	36
3.4	<i>Types</i>	37
	Numerical	37
	Categorical	38
	Free text	38
3.5	<i>Delivery formats</i>	39
	Spreadsheets	41
	Tagged formats	41
	Databases	42
	APIs	42
3.6	<i>Considerations</i>	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	<i>Planning</i>	46
3.8	<i>Conclusion</i>	47
4	Data analysis techniques	49
4.1	<i>Introduction</i>	49
4.2	<i>Close reading</i>	50



---

4.3	<i>Statistical analysis</i>	51	
	Normal distribution	52	
	Random distribution	52	
	Linear distribution	52	
	Multiple points of comparison	53	
	Techniques	53	
	Compared to machine learning	55	
4.4	<i>Machine learning</i>	55	
	Supervised learning	56	
	Unsupervised learning	56	
	Reinforcement learning	57	
4.5	<i>Natural language processing</i>	57	
	Bag-of-words	58	
	Vector analysis	59	
4.6	<i>Other tools</i>	60	
	Knowledge bases	60	
	Decision trees	60	
	Network analysis	62	
4.7	<i>Conclusion</i>	62	
5	Interpreting legal data		65
5.1	<i>Introduction</i>	65	
5.2	<i>Approaching data analysis</i>	66	
5.3	<i>Methodology</i>	67	
	Statistics	67	
	Machine learning	69	
	Natural language	69	
5.4	<i>Considerations</i>	70	
	Correlation and causation	70	
	Proxy data	71	
	Sensitivity	71	
	Suitability	72	
	Complexity	72	
	Extrapolation	73	
5.5	<i>Concerns</i>	73	
	Overfitting	74	
	Bias	75	
5.6	<i>Conclusion</i>	75	
6	Issues with using legal data		78
6.1	<i>Introduction</i>	78	



6.2	<i>Availability</i>	78	
6.3	<i>What is missing</i>	79	
6.4	<i>Ambiguity</i>	80	
6.5	<i>Limitations on language processing</i>	81	
6.6	<i>Sampling</i>	82	
	Analytics and sample size	82	
6.7	<i>Cost</i>	83	
6.8	<i>Jurisdiction</i>	84	
6.9	<i>Structure</i>	84	
	Case law	85	
	Legislation	86	
	Law firms' data	87	
6.10	<i>Risks</i>	88	
6.11	<i>Conclusion</i>	89	
7	<i>Artificial intelligence</i>		93
7.1	<i>Introduction</i>	93	
7.2	<i>Definition</i>	93	
7.3	<i>Promise and limits</i>	94	
7.4	<i>Types of artificial intelligence</i>	95	
	Statistical analysis	95	
	Machine learning	96	
	Vaporware	97	
7.5	<i>Process</i>	97	
7.6	<i>Likely impacts</i>	98	
7.7	<i>Limitations</i>	100	
7.8	<i>Concerns</i>	100	
	Fairness	100	
	Error and bias	102	
	Ethics	102	
7.9	<i>Thinking critically</i>	104	
7.10	<i>Conclusion</i>	105	
8	<i>The law and politics of legal data</i>		108
8.1	<i>Introduction</i>	108	
8.2	<i>The nature of data based applications</i>	109	
8.3	<i>Access to data</i>	110	
8.4	<i>Privacy</i>	111	
8.5	<i>Jurisdiction</i>	112	



---

Infrastructure	112	
Legacy systems	113	
Government controls	114	
8.6 <i>Primary law</i>	115	
8.7 <i>Practice data</i>	116	
8.8 <i>The research landscape</i>	116	
8.9 <i>Working toward change</i>	117	
8.10 <i>Conclusion</i>	118	
9 Vision for the future		120
9.1 <i>Introduction</i>	120	
9.2 <i>Thinking about the future</i>	121	
9.3 <i>Forecasting</i>	122	
9.4 <i>Technology adoption dynamics</i>	122	
9.5 <i>Issues</i>	124	
Data availability	124	
Regulation of technology	126	
Social concerns	127	
9.6 <i>Divergence</i>	127	
Jurisdiction	127	
Population segments	128	
Legal system	129	
Culture	129	
9.7 <i>Predictions</i>	130	
Near term	130	
Medium term	131	
Long term	132	
9.8 <i>Conclusion</i>	133	
<i>Bibliography</i>		135
<i>Index</i>		146