Contents

Acknowledgements ix

Glossary xi

1 Introduction 1

2 The development of web archiving 8
Introduction 8
Initiation: the Internet Archive 8
Diversification: the spread of web archiving 10
Collaboration: the IIPC and beyond 19
Notes and references 22

3 Selection 24
Introduction 24
The selection process 26
The selection context 28
Selection methods 29
Selection criteria 33
Defining the boundaries 36
Timing and frequency of collection 38

Web archiving tools 192
Web technology 193
Conclusion 194
Notes and references 195

Appendices

1 Web archiving and preservation tools 197
2 Model permissions form 203
3 Model test script 211
4 Model issues log 215
5 Model job description 217

Bibliography 221

Index 227

Maintenance 39
Summary 40
Notes and references 40

4 Collection methods 42
Introduction 42
The technology of the web 42
Direct transfer 46
Remote harvesting 50
Database archiving 59
Transactional archiving 61
Optimizing websites for archiving 65
Conclusions 67
Notes and references 68

5 Quality assurance and cataloguing 69
Introduction 69
Pre-collection testing 70
Post-collection testing 71
Issue tracking 73
Common issues 74
Cataloguing 78
Summary 80
Reference 81

6 Preservation 82
Introduction 82
The challenge of digital preservation 82
Preservation strategies 86
The preservation cycle 99
Passive preservation 100
Active preservation 110
Preservation metadata 120
Preserving web resources 121

Conclusion 123
Notes and references 124

7 Delivery to users 127
Introduction 127
Search and browse access 128
Placing archived content in context 131
Delivery methods 135
Functionality 141
Case study 143
Conclusion 144
Notes and references 144

8 Legal issues 146
Introduction 146
Intellectual property rights 147
Privacy 152
Content liability and human rights 154
Defending against liability 156
Enabling legislation 157
Regulatory compliance and public accountability 159
Conclusion 161
Notes and references 161

9 Managing a web archiving programme 163
Introduction 163
Alternative operational models 163
Selecting and implementing a model 170
Managing the programme 174
The UK Government Web Archive: a case study 178
Notes and references 183

10 Future trends 184
Introduction 184
Data storage 184
Digital preservation 187