

TABLE DES MATIÈRES

SOMMAIRE	5
CHAPITRE 1 Le Big data, si nous en parlions ?	9
1. Introduction	9
2. Qu'est-ce que le Big Data ?	11
2.1. Les données du Big Data	11
2.2. Les technologies et métriques du Big Data	13
2.3. Les modèles économiques du Big Data	18
3. L'écosystème français	19
3.1. L'état français vs l'état américain	19
3.2. L'Alliance Big Data	20
3.3. Les projets de Big Data en France	23
4. Les aspects éthiques du Big Data	27
Remerciements	29
Bibliographie	29
CHAPITRE 2 La magie des séries et des grands nombres : ce que l'on peut faire avec le Big Data et les archives	31
1. Introduction	31
2. Un nouveau contexte	33
3. Méthodes et approches	34
3.1. Division des tâches et des données	34
3.2. Beaucoup et pas chers !	35
3.3. Magie des grands nombres	36
3.4. Macro vs Micro analyse	38
4. Impact sur le métier	39
4.1. Nouveaux besoins et nouvelles attentes	39
4.2. Les compétences	40
Bibliographie	42

CHAPITRE 3	Le traitement automatique des langues face aux données textuelles volumineuses et potentiellement dégradées : qu'est-ce que cela change ?	43
1.	Introduction	43
2.	Les données textuelles : types et caractéristiques	44
	2.1. Types de données textuelles	45
	2.2. Caractéristiques des données textuelles	46
3.	Représentations des données textuelles	47
4.	Faire face au déluge de données	49
	4.1. Résumé automatique mono ou multi-documents	50
	4.2. Structuration et navigation	51
	4.3. Fouille d'opinion	52
	4.4. Text analytics	53
5.	Exploiter la profusion et la redondance	54
	5.1. Traduction automatique	54
	5.2. Journalisme de données	56
	5.3. Fact-checking	57
6.	Conclusion	58
	Bibliographie	59
CHAPITRE 4	Analyse automatique de l'image et du son pour l'enrichissement de contenus audiovisuels	61
1.	Introduction	61
2.	Représentation et identification	62
3.	Segmentation	63
4.	Structuration	64
5.	Transcription de la parole	65
6.	Analyse des locuteurs	68
7.	Exemple d'utilisation de l'analyse audio pour un corpus de talk-shows	69
8.	Recherche par l'exemple	70
9.	Analyse de visages	71
10.	Autres caractérisations	71
11.	Collecte des références	72
12.	Fouille de données multimédias	73
13.	Conclusion	75
	Remerciements	75
	Bibliographie	76

CHAPITRE 5	Déluge des données de la recherche ?	77
1.	Introduction	77
2.	L'évolution de la pratique scientifique	78
	2.1. Le quatrième pilier	78
	2.2. E-science et données massives	79
3.	Les données en science	80
	3.1. Définir la donnée de recherche ?	80
	3.2. Les quatre vérités des données scientifiques ?	83
	3.3. Big science et Big Data	84
	3.4. Big Data et little science	86
4.	Les enjeux de la gestion et du partage des données	87
	4.1. Vers une science ouverte	87
	4.2. La donnée scientifique, infrastructure de recherche	89
	4.3. De nouveaux métiers	92
5.	Données massives, science 2.0 et challenges sociétaux	93
	Bibliographie	94
CHAPITRE 6	Des données à l'intermédiation, une révolution économique et politique	97
1.	Introduction	97
2.	De nouveaux possibles	101
	2.1. La croissance des capacités de stockage et de traitement des données	101
	2.2. Des algorithmes toujours plus puissants	102
	2.3. L'émergence de l'à-peu-près	102
	2.4. Hic et nunc	104
3.	L'intermédiation	105
	3.1. L'art de l'intermédiation	105
	3.2. Des données primaires aux données secondaires	107
	3.3. La gouvernance des algorithmes	108
4.	L'économie des plateformes	109
	4.1. Le modèle économique biface	110
	4.2. La concentration	110
5.	Un modèle disruptif	111
	5.1. Un mode d'organisation basé sur la confiance	112
	5.2. Les secteurs du savoir	113
6.	La révolution politique	113
	6.1. Surveillance et transparence	114
	6.2. Nouveaux services publics et fiscalité	115
	6.3. L'identité et le fonctionnement démocratique	116

7. Les équilibres internationaux	117
8. Conclusion	119
Bibliographie	120
CHAPITRE 7 Extension du domaine de la veille	121
1. Introduction	121
2. Une brève histoire des pratiques de veille	122
3. Big Data versus small data ?	123
3.1. De l'information grise à l'open access	123
3.2. De l'information noire aux dark data	124
3.3. Temps réel ou temps artificiel ?	125
3.4. Du monde clos à l'univers infini	125
4. La veille : le 4 ^e V du Big Data	126
4.1. Big Data et signaux faibles	126
4.2. Du côté des éditeurs de solutions	127
4.3. Exemples d'applications	127
5. À qui appartiennent les données ?	128
6. Extension du domaine de la veille	129
Bibliographie	131
CHAPITRE 8 Travail et travailleurs de la donnée	133
1. Introduction	133
2. L'oubli du travail des données	134
3. ALGOPOL, une expérimentation sociologique sur Facebook	136
3.1. Le web enregistre la société	136
3.2. Les données du web sont accessibles, donc publiques	139
3.3. L'enquête a consenti, tout va bien	142
3.4. Les données du web sont auto-suffisantes	144
3.5. Interpréter leurs données sans les internautes	146
Bibliographie	147