

Sous la direction de
Mokrane Bouzeghoub
Rémy Mosseri

Les Big Data à découvert

CNRS ÉDITIONS

Table des matières

Préface	
<i>Alain Fuchs</i>	15
Avant-Propos	
<i>Mokrane Bouzeghoub et Rémy Mosseri</i>	16
- Première partie -	
BIG DATA : ENJEUX ET DÉFIS	
1. Introduction	
<i>Rémy Mosseri et Mokrane Bouzeghoub</i>	21
2. La déferlante des données : une courte mise en perspective d'une longue histoire	
<i>Valérie Schafer</i>	22
3. La diversité des données	
<i>Nicole Bidoit-Tollu et Anne Doucet</i>	24
4. Les technologies au service du Big Data	
<i>Hervé Fanet et Marc Duranton</i>	26
5. Place et influence des Big Data dans le commerce et l'économie	
<i>Marc Vanhuele</i>	28
6. Big Data et météorologie	
<i>Philippe Dandin</i>	30
7. La géographie et les données massives	
<i>Denise Pumain</i>	32
8. Place et influence des Big Data dans la santé	
<i>Franck Lethimonnier</i>	34
9. Big Data, individu et société	
<i>Pascal Buléon</i>	36
10. L'analyse massive de données : dérives et bonnes pratiques	
<i>Serge Abiteboul et Julia Stoyanovich</i>	38
11. Les grandes initiatives favorisant l'exploitation des Big Data	
<i>Christine Collet et Pierre Gançarski</i>	40

Conception éditoriale et graphique :
Elsa Godet - www.sciencegraphique.com

© CNRS Éditions, Paris, 2016

- Deuxième partie -

DONNÉES, ACQUISITION, STOCKAGE

1. Introduction	
<i>Henri Verdier</i>	45
2. La place des données en physique	
<i>Rémy Mosseri et Hugues Chaté</i>	46
3. Les données pertinentes pour décrire le vivant	
<i>Michaël Manuel</i>	48
4. Les collections muséologiques	
<i>Pierre Joly et Régine Vignes</i>	50
5. Les données dans l'art	
<i>Philippe Codognet</i>	52
6. Les Big Data dans la littérature	
<i>Alexandre Gefen</i>	54
7. Les données multimédias	
<i>Olivier Cappé et Vincent Claveau</i>	56
8. Les données personnelles	
<i>Claire Levallois-Barth</i>	58
9. L'univers des métadonnées	
<i>Omar Boucelma</i>	60
10. Des métadonnées aux connaissances et ontologies	
<i>Marie-Christine Rousset</i>	62
11. La qualité des données : champs de recherche et enjeux	
<i>Laure Berti-Equille</i>	64
12. Les capteurs de données	
<i>Philippe Combette et Alain Foucaran</i>	66
13. L'acquisition des données : du continu au discret	
<i>Laurent Daudet</i>	68
14. Autonomie et énergie des capteurs communicants	
<i>Gilles Grimaud et Nathalie Rolland</i>	70
15. Les réseaux de capteurs	
<i>Nathalie Mitton et Laurent Clavier</i>	72
16. Quels réseaux de communication pour les Big Data ?	
<i>Walid Dabbous</i>	74
17. Stockage de l'information : de la peinture rupestre aux data centers	
<i>Marc Bocquet et Christophe Muller</i>	76
18. Préservation et archivage	
<i>Cristinel Diaconu et Marion Massol</i>	78
19. Les grands sites de calcul et de stockage	
<i>Michel Daydé et Denis Veynante</i>	80

20. Impact énergétique des Big Data

<i>Alain Cappy</i>	82
--------------------------	----

21. Impact écologique des Big Data

<i>Françoise Berthoud et Laurent Lefèvre</i>	84
--	----

- Troisième partie -

TRAITEMENT DES DONNÉES

1. Introduction

<i>Iovka Boneva et Hélène Touzet</i>	89
--	----

2. Les traces de l'activité humaine dans le numérique

<i>Mélanie Dulong de Rosnay</i>	90
---------------------------------------	----

3. De la donnée à la décision : comment réguler par les données et les algorithmes

<i>David Restrepo Amariles et Gregory Lewkowicz</i>	92
---	----

4. Algorithmes et traitement des données

<i>Christine Froidevaux et Bruno Defude</i>	94
---	----

5. Approche algébrique de la manipulation des données

<i>François Goasdoué et Virginie Thion</i>	96
--	----

6. Requêtes et langages de manipulation de données

<i>Angela Bonifati et Emmanuel Coquery</i>	98
--	----

7. Les chaînes de traitement de données et le partage d'expériences scientifiques

<i>Sarah Cohen-Boulakia et Patrick Valduriez</i>	100
--	-----

8. Performance et optimisation

<i>Ioana Manolescu-Goujot</i>	102
-------------------------------------	-----

9. L'intégration de données massives, hétérogènes et distribuées

<i>Bernd Amann et Daniela Grigori</i>	104
---	-----

10. Privacy by Design

<i>Philippe Pucheral, Alain Rallet et Célia Zolynski</i>	106
--	-----

11. La sûreté de fonctionnement des systèmes de données

<i>Philippe Rigaux</i>	108
------------------------------	-----

12. La compression des données

<i>Christophe Collet</i>	110
--------------------------------	-----

13. La convergence du calcul scientifique et de l'analyse des données

<i>François Bodin</i>	112
-----------------------------	-----

14. MapReduce pour calculer sur des données massives

<i>Gabriel Antoniu et Mohand-Saïd Hacid</i>	114
---	-----

15. Les blockchains

<i>Jean-Paul Delahaye</i>	116
---------------------------------	-----

16. Le tournant des données en traitement automatique des langues

<i>Béatrice Daille et Adeline Nazarenko</i>	118
---	-----

- Quatrième partie -

ANALYSE DE DONNÉES ET APPRENTISSAGE

1. Introduction	
<i>Philippe Besse</i>	123
2. La prédiction dans le transport	
<i>Jean-Michel Loubes et Paola Goatin</i>	124
3. Prévission de la consommation électrique	
<i>Jean-Michel Poggi</i>	126
4. Gestion des relations client et Big Data	
<i>Jacky Akoka</i>	128
5. Corpus et traitement automatique des langues	
<i>Claire Gardent</i>	130
6. Entrepôts de données et OLAP, analyse et décision dans l'entreprise	
<i>Jérôme Darmont et Patrick Marcel</i>	132
7. Sciences des données : naissance ou renaissance ?	
<i>Gérard Biau et Pascal Massart</i>	134
8. La fouille de données	
<i>Amedeo Napoli et Alexandre Termier</i>	136
9. La fouille de texte	
<i>François Yvon</i>	138
10. Statistiques sur les données, exhaustivité et échantillonnage	
<i>Camelia Goga et Anne Ruiz-Gazen</i>	140
11. La variété dans les Big Data	
<i>Françoise Fogelman-Soulié</i>	142
12. Prédiction et hasard dans les Big Data	
<i>Giuseppe Longo</i>	144
13. L'apprentissage automatique	
<i>Jean-Philippe Vert</i>	146
14. Réseaux de neurones et apprentissage profond	
<i>Alexandre Allauzen et Yann Ollivier</i>	148
15. Optimisation à grande échelle	
<i>Francis Bach</i>	150
16. Visualisation de données massives	
<i>David Auber et Michel Beaudouin-Lafon</i>	152
17. Les systèmes de recommandations en ligne	
<i>Aurélien Garivier</i>	154
18. Extraction d'informations à partir des images	
<i>Cordelia Schmid et Jean Ponce</i>	156
19. Reconnaissance de forme sur les bases de données iconographiques	
<i>Mathias Blanc</i>	158

20. Les larges corpus d'imagerie cérébrale en neurosciences

Bertrand Thirion 160

21. L'analyse géographique et sémantique des publications scientifiques

Arnaud Banos, Hadrien Commenges et Denise Pumain 162

22. L'écosystème des Big Data

Balázs Kégl..... 164

- Cinquième partie -

WEB, RÉSEAUX SOCIAUX ET RECHERCHE D'INFORMATION

Introduction

Claire Nédellec..... 169

2. Hypertexte et hypermédia : des liens entre corpus

Jean-Marie Pierrel
 170 |

3. Recherche d'information : au cœur des moteurs de recherche

Mohand Boughanem
 172 |

4. Ordonnement de documents sur le Web

Éric Gaussier
 174 |

5. Web de données liées et Web sémantique

Fabien Gandon
 176 |

6. Extraction d'informations

Fabian Suchanek
 178 |

7. Personnalisation et adaptation au contexte

Patrice Bellot et Gabriella Pasi.....
 180 |

8. Réseaux sociaux et émergence d'opinions et de mobilisation

Jean-Philippe Cointet et Patrick Gallinari.....
 182 |

9. Des réseaux sociaux aux réseaux écologiques

Cédric Sueur.....
 184 |

10. Reconstitution des données personnelles

Amélie Marian et Arnaud Sabuguet.....
 186 |

11. L'analyse à grande échelle de nos traces numériques

Matthieu Latapy et David Chavalarias
 188 |

12. Le rôle des réseaux sociaux dans le développement de la radicalisation

Yousri Marzouki.....
 190 |

13. Les sciences participatives et les sciences du numérique se complètent-elles ?

Sihem Amer-Yahia, Jean-Gabriel Ganascia et Jean-Marc Ogier.....
 192 |

14. Les enjeux du crowdsourcing

David Gross-Amblard et Romain Julliard
 194 |

15. Éducation et analyse de l'apprentissage

Anne Boyer.....
 196 |

8 |

Table des matières | 9

16. Wikipédia, une encyclopédie en communs	
<i>Alexandre Monnin</i>	198
17. Archivage du Web	
<i>Pierre Senellart</i>	200
- Sixième partie -	
DE L'INFINIMENT PETIT À L'INFINIMENT GRAND	
Introduction	
<i>Giuseppe Baldacci et Rémy Mosseri</i>	205
2. Exploitation des données en astronomie	
<i>Emmanuel Gangler et Farouk Toumani</i>	206
3. Le CERN et les Big Data	
<i>Fairouz Malek</i>	208
4. La Terre : un système d'informations	
<i>Catherine Jeandel</i>	210
5. Océan et Big Data	
<i>Guillaume Maze et Rosemary Morrow</i>	212
6. Sismologie et Big Data	
<i>Josselin Garnier et Jean-Pierre Vilotte</i>	214
7. L'usage massif des données pour les études climatiques	
<i>Jean-Louis Dufresne et Sébastien Denvil</i>	216
8. Les grandes infrastructures analytiques de recherche : l'exemple de l'Institut Laue-Langevin	
<i>Charles Simon</i>	218
9. Le Genoscope et les infrastructures nationales	
<i>Jean Weissenbach, Claude Scarpelli et Pierre LeBer</i>	220
10. Le décryptage haut débit du vivant	
<i>Dominique Joly et Denis Faure</i>	222
11. Les enjeux du séquençage haut débit pour la biologie contemporaine	
<i>Gaëlle Lelandais et Jean-Michel Camadro</i>	224
12. Comment les Big Data font évoluer les recherches en phylogénie	
<i>Éric Baptiste et Philippe Lopez</i>	226
13. Biodiversité passée et paléogénomique	
<i>Eva-Maria Geigl et Thierry Grange</i>	228
14. Génomique, Big Data et évolution humaine	
<i>Lluís Quintana-Murci</i>	230
15. Microbes et Big Data	
<i>Jean-Michel Batto et Joël Doré</i>	232
16. Big data et diversité fonctionnelle : des organismes aux écosystèmes	
<i>Éric Garnier et Nicolas Mouquet</i>	234

17. L'écologie des systèmes	
<i>Damien Eveillard, Anne Siegel et Philippe Vandenkoornhuysse</i>	236
18. <i>Tara Oceans</i> : vers une compréhension multi-échelle de l'écosystème océan	
<i>Chris Bowler et Lucie Bittner</i>	238
19. Objets connectés et embarqués sur les animaux sauvages	
<i>Céline Le Bohec et Victor Planas-Bielsa</i>	240
20. Les données au cœur de l'infrastructure expérimentale Écotrons	
<i>Olivier Ravel et Simon Chollet</i>	242

- Septième partie -

SANTÉ HUMAINE

Introduction	
<i>Pierre Corvol</i>	247
2. Objets connectés et santé : les défis de la santé de demain	
<i>Nathalie Manaud</i>	248
3. L'imagerie cérébrale sur la piste du Big Data	
<i>Jean-François Mangin</i>	250
4. Données médicales et Big Data : limites et perspectives	
<i>Jean Charlet et Rémy Choquet</i>	252
5. Apport des Big Data à la médecine de précision en cancérologie	
<i>Pierre Saintigny, Jean-Philippe Foy et Alain Puisieux</i>	254
6. Big Data et chirurgie	
<i>Jocelyne Troccaz et Philippe Cinquin</i>	256
7. Apport des données du séquençage à la recherche sur le cancer	
<i>Alain Viari et Pierre Saintigny</i>	258
8. Big Data et santé : questions éthiques	
<i>Hervé Chneiweiss</i>	260
9. De la santé de l'environnement à la santé humaine	
<i>Pascal Simonet et Anne Maître</i>	262

- Huitième partie -

INDIVIDU ET SOCIÉTÉ

1. Introduction	
<i>Sandra Laugier</i>	267
2. Scruter les sociétés, après l'« homme moyen »	
<i>Claire Lemercier</i>	268

3. Y a-t-il une « révolution Big Data » en sociologie ?	
<i>Étienne Ollion</i>	270
4. Les tweets, des capteurs humains	
<i>Sibem Amer-Yahia, Marianne Clausel et Paule-Annick Davoine</i>	272
5. Place des Big Data dans l’histoire et l’archéologie	
<i>Bertrand Jouve et Florent Hautefeuille</i>	274
6. Enjeux sociaux des Big Data	
<i>Paola Tubaro et Antonio A. Casilli</i>	276
7. Information, communication et Big Data	
<i>Dominique Boullier</i>	278
8. Journalisme et données : de l’investigation à la robotisation	
<i>Florence Le Cam et Olivier Trédan</i>	280
9. Big Data et gouvernance urbaine	
<i>Emmanuel Eveno</i>	282
10. Développement de Smart Grids grâce aux NTIC	
<i>Benoît Robyns</i>	284
11. Smarthome : de la maison câblée à la maison écologique	
<i>James Crowley et Joëlle Coutaz</i>	286
12. Entreprises et usines du futur	
<i>Alain Bernard</i>	288
13. Agriculture et Big Data	
<i>Gilles Rabatel</i>	290
14. L’apport des Big Data dans les décisions publiques	
<i>Roberto Frega et Alexis Tsoukiàs</i>	292
15. L’uberisation de l’économie	
<i>Pascal Petit</i>	294
16. Le Big Data redessine le secteur de l’assurance	
<i>Olivier Lopez</i>	296
17. Les crypto-monnaies, une réalité virtuelle ?	
<i>Jean-Guillaume Dumas et Pascal Lafourcade</i>	298
18. Monétisation et marchandisation des données personnelles	
<i>Fabrice Rochelandet</i>	300
19. Anonymisation des données et protection de la vie privée	
<i>Françoise Dupont</i>	302
20. Les données personnelles : questions juridiques et géopolitiques	
<i>Céline Castets-Renard</i>	304
21. Place du choix individuel et droit à la déconnexion	
<i>Jessica Eynard</i>	306
22. Aux origines des GAFAM	
<i>Pascal Griset</i>	308

- Neuvième partie -

CONCLUSIONS ET PERSPECTIVES

1. Introduction	
<i>Yvon Maday</i>	313
2. Quel avenir pour la recherche en Big Data ?	
<i>Mokrane Bouzeghoub, Olivier Cappé et Balázs Kégl</i>	314
3. Robotique, Intelligence Artificielle et Big Data	
<i>Raja Chatila</i>	316
4. Le pari de l’industrialisation des neurosciences	
<i>Yves Frégnac</i>	318
5. Les avancées technologiques des mémoires électroniques	
<i>Christophe Muller et Marc Bocquet</i>	320
6. L’information quantique : de la physique à l’ingénierie	
<i>Alain Aspect et Philippe Grangier</i>	322
7. Sous les données, la société et les humains	
<i>Hervé Le Crosnier</i>	324
8. L’ouverture des données scientifiques : un enjeu fondamental	
<i>Michel Bidoit, Mokrane Bouzeghoub, Claude Kirchner et Antoine Petit</i>	326
9. Outils de collaboration internationale et de mise en réseau des données	
<i>Françoise Genova</i>	328
10. Les données dans l’enseignement de l’informatique à l’école	
<i>Gilles Dowek et Jean-Marc Petit</i>	330

- Dixième partie -

COMPLÉMENTS

Glossaire	
.....	335
Contributeurs	
.....	350
Laboratoires et structures	
.....	359
Organismes et établissements	
.....	363