

La Boîte à outils

de la

Stratégie big data

Romain Rissoan |
Romain Jouin |

DUNOD

Maquette de couverture : Caroline Joubert
Photo de la boîte : © Mega Pixel
Pictos de couverture :
© bioraven-Shutterstock.com
© Eliricon from Noun Project
© I Putu Kharismayadi from Noun Project

Mise en page : Belle Page

Le pictogramme qui figure ci-contre mérite une explication. Son objet est d'alerter le lecteur sur la menace que représente pour l'avenir de l'écrit, particulièrement dans le domaine de l'édition technique et universitaire, le développement massif du photocopillage.

Le Code de la propriété intellectuelle du 1^{er} juillet 1992 interdit en effet expressément la photocopie à usage collectif sans autorisation des ayants droit. Or, cette pratique s'est généralisée dans les établissements

d'enseignement supérieur, provoquant une baisse brutale des achats de livres et de revues, au point que la possibilité même pour

les auteurs de créer des œuvres nouvelles et de les faire éditer correctement est aujourd'hui menacée. Nous rappelons donc que toute reproduction, partielle ou totale, de la présente publication est interdite sans autorisation de l'auteur, de son éditeur ou du

Centre français d'exploitation du droit de copie (CFC, 20, rue des Grands-Augustins, 75006 Paris).



© Dunod, 2018
11 rue Paul Bert, 92240 Malakoff
www.dunod.com
978-2-10-077898-0

Le Code de la propriété intellectuelle n'autorisant, aux termes de l'article L. 122-5, 2° et 3° a), d'une part, que les « copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective » et, d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration, « toute représentation ou reproduction intégrale ou partielle faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause est illicite » (art. L. 122-4).

Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait donc une contrefaçon sanctionnée par les articles L. 335-2 et suivants du Code de la propriété intellectuelle.

VOUS AUSSI, AYEZ LE RÉFLEXE

Boîte à outils

Des outils
classés par
dossiers
thématiques

5
DOSSIER


IMAGE DE ET NOTORIÉTÉ

“
Être le meilleur est bien,
car tu es le premier.
Être unique est encore mieux,
car tu es le seul.”

Wilson Kanadi

Une présentation
visuelle de chaque outil

Exercices



EXERCICE 1 : FINALEMENT SA CONCENTRATION

- Fermez les yeux, représentez-vous le chiffre 1.
- Lorsque vous le voyez clairement en pensée, effacez de votre esprit l'image du chiffre 1.
- Représentez-vous le chiffre 2. Continuez ainsi jusqu'à 10.

EXERCICE 2 : LA MÉTHODE DE « L'ÉCOUTE AVEC LE CŒUR »

> La technique se résume en cinq questions

1. Qui s'est-il passé ?

Quelle émotion avez-vous ressentie ?

Quelle a été la plus difficile pour vous ?

Outil 33 Le Personal Branding

“
Aujourd'hui,
à l'ère de l'individu,
vous devez
être votre propre
marque.”

En quelques mots

Le Personal Branding ou la gestion de sa marque personnelle est un outil de réflexion et de mise en œuvre d'actions définies visant à contribuer à la construction de son image personnelle. En marketing de soi, le Personal Branding est l'ensemble des moyens, techniques et canaux que l'on utilise afin de construire son identité, se rendre visible et se promouvoir de façon pertinente et efficace. À l'instar des entreprises qui créent des marques, les rendent visibles, développent leur notoriété et travaillent leur image. Il est possible et utile de construire et mettre en avant sa propre « marque ».

LES COMPOSANTES DE LA VALEUR DE L'EXPIÉRIENCE POUR LE CLIENT

Composants de la valeur perçue dans l'expérience	Facteurs associés par l'entreprise à l'impact de cette valeur
Émotion Fait espérer ou gîger de l'orgueil	des offres spéciales, des ventes flash, des cadeaux à gîger, des chartes ou des algorithmes géniaux...
Reconnaissance Fait gîger du temps ou respecte l'urgence du client	une aventure 24h/24, une livraison instantanée...

Des exemples,
cas ou exercices
pour approfondir



La Boîte à outils

DES OUTILS OPÉRATIONNELS TOUT DE SUITE

MEGA Boîte à Outils

Manager leader - 100 outils

Coordonnée par Pascale Bêlorgey,
Nathalie Van Laethem

Digital - 100 outils

Coordonnée par Catherine Lejealle

MÉTIERS

Acheteur, 3^e éd.

Stéphane Canonne, Philippe Petit

Assistante, 2^e éd.

Christine Harache, Héliène Tellitocci

Auditeur financier, 2^e éd.

Sylvan Boccon-Gibod, Eric Vilmint

Chef de produit, 2^e éd.

Nathalie Van Laethem, Stéphanie Moran

Chef de projet, 2^e éd.

Jérôme Maes, François Debois

Coach en entreprise, 2^e éd.

Belkacem Ammiar, Omid Kohneh-Chahri

Commercial, 3^e éd.

Pascale Bêlorgey, Stéphane Mercier

Community Manager

Clément Pellerin

Comptabilité, 2^e éd.

Bruno Bachy

Consultant, 2^e éd.

Patrice Stern, Jean-Marc Schoettl

Contrôle de gestion

Caroline Selmer

Création d'entreprise, 2018

Catherine Léger-Jarniou,
Georges Kalousis

E-commerce

Christian Delabre

Formateurs, 3^e éd.

Fabienne Bouchut, Isabelle Cauden,
Frédérique Cuisiniez

Management, 2^e éd.

Patrice Stern, Jean-Marc Schoettl

Micro-entrepreneur

Jacques Hellart, Caroline Selmer

Pilote des systèmes d'information, 2^e éd.

Jean-Louis Foucard

Publicité

Servanne Barre, Anne-Marie
Gayraud-Carrera

Responsable communication, 3^e éd.

Bernadette Jézéquel, Philippe Gérard

Responsable financier, 2^e éd.

Caroline Selmer

Responsable marketing omnicanal, 3^e éd.

Nathalie Van Laethem, Béatrice Durand-
Mégret

Responsable qualité, 3^e éd.

Florence Gillet-Goinard, Bernard Seno

Ressources humaines, 2^e éd.

Annick Haegel

Santé - Sécurité - Environnement, 3^e éd.

Florence Gillet-Goinard, Christel Monar

TPE

Guillaume Ducret

COMPÉTENCES TRANSVERSALES

Conduite du changement

David Autissier, Jean-Michel Moutot

Créativité, 2^e éd.

François Debois, Arnaud Groff,
Emmanuel Chenevier

Design management

Bérangère Szostak, François Lenfant

Développement durable et RSE

Vincent Maymo, Geoffroy Murat

Gestion des conflits

Jacques Salzer, Arnaud Stimec

Innovation, 2^e éd.

Géraldine Benoit-Cervantes

Intelligence collective

Béatrice Arnaud, Sylvie Caruso-Cahn

Intelligence économique

Christophe Deschamps, Nicolas Moinet

Lean

Radu Demetrescoux

Leadership, 2^e éd.

Jean-Pierre Testa, Jérôme Lafargue,
Virginie Tilhet-Coartet

Management de la relation client, 2^e éd.

Laurence Chabry, Florence Gillet-Goinard,
Raphaëlle Jourdan

Management transversal

Jean-Pierre Testa, Bertrand Déroulède

Marketing digital

Stéphane Truphème, Philippe Gastaud

Mind mapping

Xavier Delengaïne, Marie-Rose
Delengaïne

Mon parcours professionnel

Florence Gillet-Goinard, Bernard Seno

Négociation, 2^e éd.

Patrice Stern, Jean Mouton

Organisation, 2^e éd.

Benoît Pommeret

Prise de décision

Jean-Marc Santi, Stéphane Mercier,
Olivier Arnould

Réseaux sociaux, 4^e éd.

Cyril Bladier

Sécurité économique

Nicolas Moinet

Stratégie, 2^e éd.

Bertrand Giboin

Stratégie digitale omnicanale

Catherine Headley, Catherine Lejealle

Supply chain

Alain Perrot, Philippe Villemus

DÉVELOPPEMENT PERSONNEL**Bien-être au travail**

Clothilde Huet, Gaëlle Rohou,
Laurence Thomas

Confiance en soi

Annie Leibovitz

**Développement personnel
en entreprise**

Laurent Lagarde

Efficacité professionnelle

Pascale Bêlorgey

Gestion du stress

Gaëlle du Penhoat

Gestion du temps

Pascale Bêlorgey

Intelligence émotionnelle

Martine-Eva Launet, Céline Peres-Court

Marketing de soi

Nathalie Van Laethem, Stéphanie Moran

Motivation

Sophie Micheau-Thomazeau,
Laurence Thomas

Pleine conscience au travail

Sylvie Labouesse, Nathalie Van Laethem

Avant-propos

“

Un bon data scientist est intéressé par la résolution de problèmes, pas par de nouveaux outils.

KDNuggets

L'ouvrage que nous vous proposons est dédié au domaine de la data. Il a pour ambition de dresser un panorama des notions, concepts et outils les plus éprouvés sur le marché actuellement. Sans être technique, il vous permettra de connaître les cas d'usage des différents outils et comment ils se distinguent les uns des autres.

Le monde de la data

En 2012, la *Harvard Business Review* (*HBR*) publiait un article : « Data-scientist, the sexiest job of the 21st century ». Depuis, tout le monde veut être data scientist : le gouvernement français a fait de la data science un des piliers de la « nouvelle France industrielle » en 2015, la BPI finance à tour de bras des start-up proposant du machine learning ou du big data et la récente étude « France IA » montre à quel point la France a une carte à jouer dans ce nouveau domaine de l'analyse des données, mais que la partie n'est pas gagnée.

Ce que la *HBR* avait oublié de nous dire, c'est que data scientist est un travail compliqué. Un data scientist doit avoir 3 compétences :

1. informatique
2. mathématique
3. business

Complétées par une excellente capacité à communiquer, c'est-à-dire transmettre ces trois compétences majeures à un auditoire qui ne les a pas.

Ce contexte rend la tâche du data scientist quasi impossible à atteindre. C'est pourquoi nous travaillons en équipe, avec chacun ses compétences. Avec ce livre, vous apprendrez le vocabulaire nécessaire au travail en équipe indispensable dans le monde de la data. Vous pourrez ainsi mettre vos compétences particulières à disposition de vos collègues.

Ce livre est avant tout un livre sur la « stratégie big data » : aussi, ses quatre premiers dossiers s'attachent à expliciter les enjeux qui se cachent derrière ce phénomène :

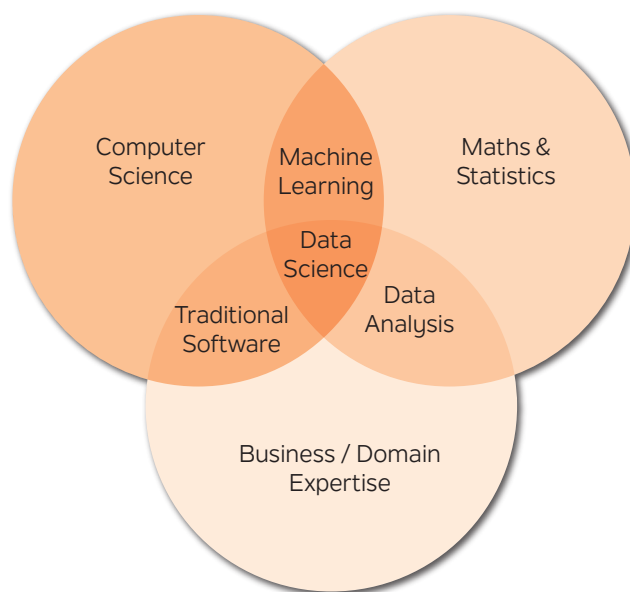
- Qu'est-ce qui a changé ces dernières années vis-à-vis de la data ?
- Quels sont les différents types de data ?
- Comment en tirer parti pour la stratégie de son entreprise ?
- Quels impacts techniques ?

Ces dossiers intéresseront plutôt le manager avide de comprendre comment profiter de cette opportunité qu'est le big data.

Viennent ensuite trois dossiers plus opérationnels, allant de la gestion de projets data à l'introduction des concepts de machine learning, et la différence avec les concepts de big data. Les personnes récemment nommés chef de projets dans un environnement big data y trouveront les clefs pour communiquer et comprendre les équipes avec lesquelles elles devront travailler.

Le dernier dossier, sur le passage en production, est une introduction aux différentes technologies usuellement mises en place dans les infrastructures big data. Il permet d'en comprendre les concepts, sans rentrer dans le détail technique opérationnel en tant que tel. Ainsi, les différents chapitres sont relativement indépendants, et s'adressent à un public de managers, ou chefs de projet, curieux du phénomène « big data ».

Classiquement, la data science nécessite six expertises distinctes, que l'on cherche chez le data scientist. Évidemment, il est bien difficile de disposer de l'ensemble, c'est pourquoi les projets sont généralement effectués en équipe. Il faut alors s'assurer de partager le même vocabulaire pour que le projet avance correctement.



Les compétences des data scientists

Sommaire

	Avant-propos	6
DOSSIER 1	L'ÈRE DE LA DATA	10
	• <i>Outil 1</i> La data (la donnée)	12
	• <i>Outil 2</i> Le sens de la donnée (donnée brute)	14
	• <i>Outil 3</i> Deep Learning, machine learning, intelligence artificielle	18
	• <i>Outil 4</i> Le V du sens de la donnée	20
	• <i>Outil 5</i> Le V de la diversité	24
	• <i>Outil 6</i> Le V de la puissance des données	28
	• <i>Outil 7</i> Vie privée et big data	32
DOSSIER 2	ENJEUX ORGANISATIONNELS DU BIG DATA	36
	• <i>Outil 8</i> Smart data, fast data, big data	38
	• <i>Outil 9</i> Les métiers du big data	40
	• <i>Outil 10</i> Informatisation et modernisation des systèmes informatiques	42
	• <i>Outil 11</i> Big data vs Business Intelligence au service de la marque employeur	44
	• <i>Outil 12</i> Communication transparente et spontanée	46
	• <i>Outil 13</i> L'ouverture et l'intégration : Open Source, SAAS et Webservice	48
	• <i>Outil 14</i> EDM, MDM, DMP et ETL	50
	• <i>Outil 15</i> S'appuyer sur le plan de l'entreprise	52
	• <i>Outil 16</i> S'appuyer sur les objectifs annuels	54
	• <i>Outil 17</i> S'appuyer sur des ateliers d'idéation	56
DOSSIER 3	ENJEUX STRATÉGIQUES DU BIG DATA	58
	• <i>Outil 18</i> Analyse de l'environnement : Deeplist	60
	• <i>Outil 19</i> Analyse des enjeux big data : le SWOT	64
	• <i>Outil 20</i> Influenceurs du big data : les parties prenantes	68
	• <i>Outil 21</i> Appétence et maturité : le Cycle de vie	72
	• <i>Outil 22</i> Expectatives : la matrice BCG	76
	• <i>Outil 23</i> Intensité concurrentielle : la matrice Porter	78
	• <i>Outil 24</i> Digitalisation des services : la Chaîne de valeur	80
	• <i>Outil 25</i> Les risques du big data	82
	• <i>Outil 26</i> Choisir de devenir Data Driven	84
DOSSIER 4	ENJEUX TECHNIQUES DU BIG DATA	86
	• <i>Outil 27</i> Données structurées et non structurées	88
	• <i>Outil 28</i> Lac de données	90
	• <i>Outil 29</i> Stockage distribué	92
	• <i>Outil 30</i> Calcul distribué	94
	• <i>Outil 31</i> Théorème de CAP	96

DOSSIER 5	GESTION DU PROJET	98
	• <i>Outil 32</i> Gérer une équipe.....	100
	• <i>Outil 33</i> Formuler une question.....	102
	• <i>Outil 34</i> Comprendre les données.....	104
	• <i>Outil 35</i> Visualiser les données : la data visualisation.....	106
	• <i>Outil 36</i> Data cleaning : que faire des valeurs aberrantes ?.....	108
	• <i>Outil 37</i> Tester des modélisations.....	110
	• <i>Outil 38</i> Performance d'une classification binaire.....	112
	• <i>Outil 39</i> Performance d'une régression.....	116
	• <i>Outil 40</i> Le storytelling.....	118
	• <i>Outil 41</i> Présenter des résultats actionnables.....	120
	• <i>Outil 42</i> Enrichir le data set.....	122
	• <i>Outil 43</i> Agilité et Scrum.....	124
DOSSIER 6	USAGE ET MAÎTRISE DES ALGORITHMES	126
	• <i>Outil 44</i> Typologies des algorithmes de machine learning.....	128
	• <i>Outil 45</i> Principes des algorithmes d'apprentissage supervisé.....	130
	• <i>Outil 46</i> Principes des algorithmes d'apprentissage non supervisé.....	132
	• <i>Outil 47</i> Principe des algorithmes par renforcement.....	134
	• <i>Outil 48</i> Principes du Deep Learning.....	136
	• <i>Outil 49</i> Les arbres de décision.....	138
	• <i>Outil 50</i> Le couteau suisse du data scientist : le Random Forest.....	140
DOSSIER 7	CHOIX DES TECHNOLOGIES BIG DATA	142
	• <i>Outil 51</i> Python.....	144
	• <i>Outil 52</i> R : statistiques, cran, ggplot.....	148
	• <i>Outil 53</i> Scala et la programmation fonctionnelle.....	150
	• <i>Outil 54</i> Plateforme sur le Web : Dataiku.....	152
	• <i>Outil 55</i> L'écosystème Hadoop et la distribution Hortonworks.....	154
	• <i>Outil 56</i> Spark vs Flink.....	158
	• <i>Outil 57</i> SMACK : Spark, Mesos, Akka, Cassandra, Kafka.....	162
	• <i>Outil 58</i> Lambda et Kappa architecture.....	166
	• <i>Outil 59</i> Aller dans le cloud ?.....	168
DOSSIER 8	MISE EN PRODUCTION	170
	• <i>Outil 60</i> Penser l'infrastructure.....	172
	• <i>Outil 61</i> DevOps.....	174
	• <i>Outil 62</i> Docker.....	176
	• <i>Outil 63</i> Infrastructure as code.....	178
	• <i>Outil 64</i> Haute disponibilité et Redondance.....	180
	• <i>Outil 65</i> Mise à jour des modèles prédictifs.....	182
	Glossaire.....	185
	Crédits iconographiques.....	191