

SpringerBriefs in Computer Science

Series Editors

Stan Zdonik, Brown University, Providence, RI, USA

Shashi Shekhar, University of Minnesota, Minneapolis, MN, USA

Xindong Wu, University of Vermont, Burlington, VT, USA

Lakhmi C. Jain, University of South Australia, Adelaide, SA, Australia

David Padua, University of Illinois Urbana-Champaign, Urbana, IL, USA

Xuemin Sherman Shen, University of Waterloo, Waterloo, ON, Canada

Borko Furht, Florida Atlantic University, Boca Raton, FL, USA

V. S. Subrahmanian, Department of Computer Science, University of Maryland,
College Park, MD, USA

Martial Hebert, Carnegie Mellon University, Pittsburgh, PA, USA

Katsushi Ikeuchi, Meguro-ku, University of Tokyo, Tokyo, Japan

Bruno Siciliano, Dipartimento di Ingegneria Elettrica e delle Tecnologie
dell'Informazione, Università di Napoli Federico II, Napoli, Italy

Sushil Jajodia, George Mason University, Fairfax, VA, USA

Newton Lee, Institute for Education, Research and Scholarships,
Los Angeles, CA, USA

SpringerBriefs present concise summaries of cutting-edge research and practical applications across a wide spectrum of fields. Featuring compact volumes of 50 to 125 pages, the series covers a range of content from professional to academic.

Typical topics might include:

- A timely report of state-of-the art analytical techniques
- A bridge between new research results, as published in journal articles, and a contextual literature review
- A snapshot of a hot or emerging topic
- An in-depth case study or clinical example
- A presentation of core concepts that students must understand in order to make independent contributions

Briefs allow authors to present their ideas and readers to absorb them with minimal time investment. Briefs will be published as part of Springer's eBook collection, with millions of users worldwide. In addition, Briefs will be available for individual print and electronic purchase. Briefs are characterized by fast, global electronic dissemination, standard publishing contracts, easy-to-use manuscript preparation and formatting guidelines, and expedited production schedules. We aim for publication 8–12 weeks after acceptance. Both solicited and unsolicited manuscripts are considered for publication in this series.

More information about this series at <http://www.springer.com/series/10028>

Renata Rawlings-Goss

Data Science Careers, Training, and Hiring

A Comprehensive Guide to the Data
Ecosystem: How to Build a Successful
Data Science Career, Program, or Unit

Renata Rawlings-Goss
Georgia Institute of Technology
Atlanta, GA, USA

ISSN 2191-5768 ISSN 2191-5776 (electronic)
SpringerBriefs in Computer Science
ISBN 978-3-030-22409-7 ISBN 978-3-030-22407-3 (eBook)
<https://doi.org/10.1007/978-3-030-22407-3>

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2019

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

*To Malcolm and Marcus for whom all things
are possible*

Acknowledgments

I would like to deeply thank Kendra Lewis-Strickland, Illona Sheffey, Stanford Goss, and Keith Rawlings for their unwavering support in preparing this manuscript. Thank you for diligently searching through references, editing, and genuinely strengthening this work with your unique gifts.

Contents

- 1 Introduction 1**
 - 1.1 The Data Science Opportunity 1
 - 1.2 Data People: An Extended Metaphor. 2
 - References. 3
- 2 Building Data Careers 5**
 - 2.1 Assessing Data-Enabled Career Paths 6
 - 2.1.1 The Data Generalist 7
 - 2.2 The Rise of the Data Specialists 9
 - 2.2.1 Artificial Intelligence: Machines Acting Like People 9
 - 2.2.2 Machine Learning (ML): Learning from the Past 11
 - 2.2.3 Computer Vision and Image Processing: Recognizing Objects and Images 12
 - 2.2.4 Natural Language Processing (NLP): Understanding Written Language. 13
 - 2.2.5 Internet of Things (IoT): Sensors Bring the Internet into the Real World 14
 - 2.3 Skills for the Career You Want. 14
 - 2.3.1 Innovation Skills (Not “Soft Skills”). 15
 - 2.3.2 Four Invaluable Skills for Success. 15
 - 2.3.3 Technical Skills 18
 - 2.4 Learning Data Science 19
 - 2.4.1 Evaluating and Finding the Perfect Pathway 19
 - 2.4.2 Picking a Career Path. 23
 - 2.5 The Job Market 24
 - 2.5.1 Size and Scale 24
 - 2.5.2 Skills in Demand 25
 - 2.5.3 Data Career Experiences 25
 - 2.5.4 Location Matters 26
 - 2.5.5 Who’s Hiring 27
 - References. 28

3	Building Data Programs	31
3.1	A GPS for Learning and Work	31
3.2	Institutional Culture	32
3.3	Interdisciplinary Collaborations (The Role of Faculty)	33
3.4	U.S. Models for Data-Focused Programs	34
3.4.1	Building a New Academic Unit	34
3.4.2	Expanding an Existing Unit	35
3.4.3	Data Literacy for All	35
3.4.4	Creating New Connectors	36
3.4.5	Creating New Stand-Alone Entity	37
3.4.6	Data Residency or Exchange Program	37
3.4.7	Career Services and Support	38
3.5	Resources for Data Science Curriculum	38
3.5.1	Data Science Module Curriculum: Learning Levels	39
3.6	Continuous Instructor Learning	40
3.6.1	Faculty Career Advancement	40
3.6.2	Benefits of Interdisciplinary Collaborations	41
3.6.3	Faculty Training and Credentialing	41
3.6.4	Faculty Recruitment	42
3.6.5	Collaborations with Industry and Government	42
3.7	Access to Data	43
3.7.1	Trusting Data	44
3.7.2	Cleaning Data	45
3.7.3	Resources for Data	45
3.7.4	Data Program Solutions	46
3.8	Top Recommendations	47
	References	49
4	Building Data Talent and Workforce	53
4.1	Why Is Hiring for Data Science and Analytics Different?	53
4.2	Don't Go It Alone: Build a Culture of Recruiting	54
4.2.1	The Traditional Role of Human Resources	54
4.2.2	An Updated Culture of Recruiting	55
4.2.3	The Foundation of Partnership	57
4.2.4	The Missing Link	58
4.3	Reasonable Expectations	58
4.3.1	The Data C-Suite	59
4.4	Skills Assessment	60
4.5	Talent Sourcing: An Eye Toward Diversity	61
4.5.1	Academia	62
4.5.2	Startups	64
4.5.3	Industry Partners, Consultants, and Non-Traditional Partners	65
4.6	Continuing Education for Managers and Workers	65
	References	66

5 Conclusion 67

5.1 The New World 68

6 Resources 69

6.1 Program Overview: Over 450 Data Science Degree Programs
Across the Nation. 69

About the Author

Renata Rawlings-Goss is a Data Strategic Coach and Author who helps professionals and executives cut through the clutter of the Data Science landscape to find their niche.

She is the current Executive Director of the South Big Data Regional Innovation Hub, one of only four federally funded Big Data Innovation Hubs for the nation, serving 16 states—Delaware through Texas—in forming Data Science partnerships between industry, academia, and government. She is also the Director of Industry Partnerships for the Institute of Data Engineering and Science at the Georgia Institute of Technology (Georgia Tech).

Previously, Dr. Rawlings-Goss worked with the White House Office of Science and Technology Policy, under President Obama, to create the National Data Science Organizers group. She also co-led the writing team for the Federal Big Data Strategic Plan, including 19 federal agencies, and was awarded an AAAS Big Data Science and Technology Policy Fellowship with the National Science Foundation in the directorate of Computer and Information Science and Engineering (CISE-OAD).

Dr. Rawlings-Goss lives in Atlanta, GA, and received her training in computational genomics, biophysics, and physics from the University of Pennsylvania, the University of Michigan, and Florida A&M University, respectively. Her research interests included data-driven analysis of African and African-American genetic expression as well as professional interests in Data Science education and workforce development in communities of color.

Abbreviations

AI	Artificial intelligence
HR	Human resources
IoT	Internet of Things
ML	Machine learning
NLP	Natural language processing

About the Book

This book is an information-packed overview of how to structure a Data Science career, a Data Science degree program, and how to hire a Data Science team, including resources and insights from the author's experience with national and international large-scale data projects as well as industry, academic and government partnerships, education, and workforce.

Outlined here are tips and insights into navigating the data ecosystem as it currently stands, including career skills, current training programs, as well as practical hiring help and resources. Also, threaded through the book is the outline of a data ecosystem, as it could ultimately emerge, and how career seekers, training programs, and hiring managers can steer their careers, degree programs, and organizations to align with the broader future of Data Science. Instead of riding the current wave, this book ultimately seeks to help professionals, programs, and organizations alike prepare a sustainable plan for growth in this ever-changing world of data.

The book is divided into three sections, the first, "Building Data Careers," addresses the perspective of a potential career seeker interested in a career in data; the second, "Building Data Programs," is from the perspective of a newly formed Data Science degree or training program; and the third, "Building Data Talent and Workforce," is from the perspective of a Data and Analytics Hiring Manager. Each is a detailed introduction to the topic with practical steps and professional recommendations.

The reason for presenting the book from different points of view is that in the fast-paced data landscape, it is helpful to each group to more thoroughly understand the desires and challenges of the other. It will, for example, help the career seekers to understand best practices for hiring managers to better position themselves for jobs. It will be invaluable for data training programs to gain the perspective of career seekers who they want to help and attract as students. Also, hiring managers will not only need data talent to hire but also workforce pipelines that can only come from partnerships with the universities, data training programs, and educational experts. The interplay gives a broader perspective from which to build.