

New Horizons for a Data-Driven Economy

José María Cavanillas • Edward Curry •
Wolfgang Wahlster
Editors

New Horizons for a Data-Driven Economy

A Roadmap for Usage and Exploitation of
Big Data in Europe

 Springer Open

Editors

José María Cavanillas
Atos Spain, S.A.
Madrid, Spain

Edward Curry
National University of Ireland Galway
Galway, Ireland

Wolfgang Wahlster
German Research Centre for
Artificial Intelligence (DFKI)
Saarbrücken, Germany

ISBN 978-3-319-21568-6

ISBN 978-3-319-21569-3 (eBook)

DOI 10.1007/978-3-319-21569-3

Library of Congress Control Number: 2015951834

© The Editor(s) (if applicable) and The Author(s) 2016. The book is published open access at SpringerLink.com.

Open Access This book is distributed under the terms of the Creative Commons Attribution-Noncommercial 2.5 License (<http://creativecommons.org/licenses/by-nc/2.5/>) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

The images or other third party material in this book are included in the work's Creative Commons license, unless indicated otherwise in the credit line; if such material is not included in the work's Creative Commons license and the respective action is not permitted by statutory regulation, users will need to obtain permission from the license holder to duplicate, adapt, or reproduce the material.

This work is subject to copyright. All commercial 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.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

Foreword¹

This book reports on preparatory work toward an important policy objective of the European Commission: turning Europe into a safe and privacy-respecting society that thrives by extracting maximum value from the data it produces and reuses, be it in support of important societal goals or as fuel for innovation in productive activities.

Our plans for Europe are described in our July 2014 Communication on a data-driven economy, where we spell out a three-pronged approach addressing regulatory issues (such as personal data protection and data ownership), framework conditions (such as data standards and infrastructures), and community building.

The first visible step of our community building efforts is a massive commitment (534 million Euros by 2020), which we signed in October 2014, to enter in a Public Private Partnership with the Big Data Value Association (BDVA): with the help from industrial parties and groups that represent relevant societal concerns (such as privacy), we intend to identify and solve technical problems and framework conditions (such as skill development) that stand in the way of European companies increasing their productivity and innovativeness by making efficient use of data technologies. By shouldering some of the financial risk of these activities, we plan to leverage even more massive European investment: for every public Euro invested by the Commission, our industry partners have committed to investing four private Euros.

Naturally, this requires some well-informed and clear thinking on which domains of data-related activities hold the greatest promise for a safe and prosperous Europe and on how we can avoid wasteful duplication in the development of data infrastructures, formats, and technologies. The book you are holding in your hands gives you a first lay of the land: it results from more than two years of work

¹ The views expressed in the article are the sole responsibility of the author and in no way represent the view of the European Commission and its services.

(also funded by the European Union) aimed at identifying issues and opportunities that are specifically European in character.

We fully expect that many of these results will be included and further elaborated over the years in the strategic planning of the BDVA, and we are happy to share them in this book with the broader public.

We hope that you will find them informative and that they will help you shape your own thinking on what your expectations and active role might be in a better Europe that has taught itself to run on data.

Luxembourg City, Luxembourg
October 2015

Giuseppe Abbamonte
Directorate G Media & Data
European Commission DG CONNECT

Foreword

Data has become a factor just as important to production as labor, capital, and land. For the new value creators in today's technology start-ups, little capital and office space is required. Both can be almost free when a firm is growing 1 % per day, on any metric. But without talent, and without the right kind of data, such a takeoff is highly improbable.

We see the same forces at play in SAP's Innovation Center Network. Attracting the right talent was critical to establish the first Innovation Center in Potsdam. And having large, real-world datasets from customers and co-innovation partners is critical to many of our innovations. To make a difference in cancer treatment and research with our Medical Research Insights app, we critically depended on data-driven collaboration with the National Center for Tumor Diseases. The same holds for incubating SAP's new sports line of business by co-innovating with the German national soccer team based on real-time sensor feeds from their players. And it holds true for SAP's many initiatives in the Internet of Things, like the predictive maintenance apps with John Deere and Kaeser.

The Big Data Value Association (BDVA) is poised to make a difference both for data availability and for talent. By bringing together businesses with leading researchers, software and hardware partners, and enabling co-innovation around large, real-world datasets, BDVA can help lower the data barrier. And helping educate the next generation of thought leaders, especially in data science, computer science, and related fields, BDVA can help increase the supply of talent. Both are critical so Europe can begin to lead, not follow, in creating value from big data.

By clearly defining the opportunity in big data, by examining the big data value chain, and by deep-diving into industry sector applications, this book charts a way forward to new value creation and new opportunities from big data. Decision makers, policy advisors, researchers, and practitioners on all levels can benefit from this.

Berlin, Germany
Brussels, Belgium

Jürgen Müller
Vice President, SAP Innovation Center Network
President, Big Data Value Association

Preface

Welcome to our humble contribution to the huge universe of big data literature. We could ironically say there are almost as many books, leaflets, conferences, and essays about the possibilities of big data as data itself to be collected, curated, stored, and analyzed, yet a single zettabyte of useful data is an amount of information we are currently incapable of writing, and as described in this book, 16 zettabytes of data are waiting for us in 2020.

However, according to many research and industrial organisations, this contribution is actually not that humble and is even unique in many senses.

First of all, this book is not just another approach made by a single player looking down from a corner of the world. It is the compendium of more than 2 years of work performed by a set of major European research centers and industries. It is the compilation and processed synthesis of what we all have done, prepared, foreseen, and anticipated in many aspects of this challenging technological context that is becoming the major axis of the new digitally transformed business environment.

But the most important part of the book is you, the reader. It is commonly said that “a map is useless for the one who does not know where to go.” This book is a map. An immediate goal of this book is to become a “User’s Manual” for those who want to blaze their own trail in the big data jungle. But it can also be used as a reference book for those experts who are sailing their own big data ship and want to clarify specific aspects on their journey.

You reader, either trailblazer or old sailor, have to make your own way through the book. In this map, you will not only find answers and discussions about legal aspects of big data but also about social impact and education needs and requirements. You will also find business perspectives, discussions, and estimations of big data actuations in the different sectors of the economy, ranging from the public sector to the retailing actors. And you will also find technological discussions about the different stages of data and how to address these emerging technologies.

We worked on all these matters within the context of a European Commission project called BIG (Big Data Public Private Forum), which was an enormous challenge and one that we reckon has been successfully achieved and accomplished.

The book is divided into four parts: Part I “The Big Data Opportunity” explores the value potential of big data with a particular focus on the European context. Chapter 1 sets the scene for the value potential of big data and examines the legal, business, and social dimensions that need to be addressed to deliver on its promise. Next, Chap. 2 briefly introduces the European Commission’s BIG project and its remit to establish a big data research roadmap for Horizon 2020 to support and foster research and innovation in the European Research Area.

Part II “The Big Data Value Chain” details the complete big data lifecycle from a technical point of view, ranging from data acquisition, analysis, curation, and storage to data usage and exploitation. Chapter 3 introduces the core concepts of the big data value chain. The next five chapters detail each stage of the data value chain, including a state-of-the-art summary, emerging use cases, and key open research questions. Chapter 4 provides comprehensive coverage of big data acquisition, which is the process of gathering, filtering, and cleaning data before it is put in a data warehouse or any other storage solution for further processing. Chapter 5 discusses big data analysis that focuses on transforming raw acquired data into a coherent, usable resource suitable for analysis to support decision-making and domain-specific usage scenarios. Chapter 6 investigates how the emerging big data landscape is defining new requirements for data curation infrastructures and how big data curation infrastructures are evolving to meet these challenges. Chapter 7 provides a concise overview of big data storage systems that are capable of dealing with high velocity, high volumes, and high varieties of data. Finally, Chap. 8 examines the business goals that need access to data and their analyses and integration into business decision-making in different sectors.

Part III “Usage and Exploitation of Big Data” illustrates the value creation possibilities of big data applications in various sectors, including industry, healthcare, finance, energy, media, and public services. Chapter 9 provides the conceptual background and overview of big data-driven innovation in society, highlighting factors and challenges associated with the adequate diffusion, uptake, and sustainability of big data-driven initiatives. The remaining chapters describe the state of the art of big data in different sectors, examining enabling factors, industrial needs, and application scenarios and distilling the analysis into a comprehensive set of requirements across the entire big data value chain. Chapter 10 details the wide variety of opportunities for big data technologies to improve overall healthcare delivery. Chapter 11 investigates the potential value to be gained from big data by government organizations by boosting productivity in an environment with significant budgetary constraints. Chapter 12 explores the numerous advantages of big data for financial institutions. Chapter 13 examines the domain-specific big data technologies needed for cyber-physical energy and transport systems, where the focus needs to move beyond big data to smart data technologies. Chapter 14 discusses the media and entertainment sectors which are in many respects an early adopter of big data technologies because it enables them to drive digital transformation, exploiting more fully not only data which was already available but also new sources of data from both inside and outside the organization.

Finally, Part IV “A Roadmap for Big Data Research” identifies and prioritizes the cross-sectorial requirements for big data research and outlines the most urgent and challenging technological, economic, political, and societal issues for big data in Europe. Chapter 15 details the process used to consolidate the big data requirements from different sectors into a single prioritized set of cross-sector requirements that were used to define the technology policy, business, and society roadmaps together with action recommendations. Chapter 16 describes the roadmaps in the areas of technology, business, policy, and society. The chapter introduces the Big Data Value Association (BDVA) and the Big Data Value contractual Public Private Partnership (BDV cPPP) which provide a framework for industrial leadership, investment, and commitment of both the private and public sides to build a data-driven economy across Europe.

We invite you to read this book at your convenience, and we wish that you will enjoy it as much as we have whilst preparing its contents.

Ciudad Real, Spain
Galway, Ireland
Saarbrücken, Germany
October 2015

José María Cavanillas
Edward Curry
Wolfgang Wahlster

Book Acknowledgements

The authors would like to acknowledge the support, openness, and good humor of all their colleagues who contributed to the BIG project and to this book in ways both big and small. The editors would like to thank Meggan Gregory for her support in the preparation of the final manuscript. Thanks also go to Ralf Gerstner, Viktoria Meyer, and all at Springer for their professionalism and assistance throughout the journey of this book.

Project Acknowledgements

The first person I would like to acknowledge is Edward Curry, the real alma mater of this book. He and his team at the National University of Ireland Galway have been working hard for the BIG project and fueling strongly this compilation from their rainy, green, and blue beautiful hometown where we met, under the sunshine, on that very first day. That day he promised us all that it was the only sunny day in Galway that year. He brought slides of the Galway streets under the rain to show he was telling the truth, and I was even accused of having brought the sun from Spain.

Many people have collaborated for this success as well in our company Atos. Our Atos colleagues, Alicia García, Blanca Jordán, Nuria de Lama, Felicia Lobillo, Josep Martrat, Ricard Munné, Tomás Pariente, Ana María Piñuela, Pedro Soria, and too many others to be mentioned, have worked hard and participated in this success. Let me mention as well the initial leadership of Diana Pottecher, who spun off to carry out her own data business in the wilderness of the Internet.

In the case of the Deutsche Forschungszentrum für Künstliche Intelligenz, DFKI, the impulse coming from its CEO, Wolfgang Wahlster, was a leveraging factor for receiving the work of researchers such as Tilman Becker, Sabrina Hoppe, and Denise Paradowski as well as many others.

The participation of Siemens has been especially welcome, namely from researchers Stephan Guido, Yi Huang, Johannes Riedl, Sebnem Rusitschka, and Sonja Zillner, as well as from the large team at Siemens Research.

The French company Exalead has been one of the main fuellers of the project and its contents, mainly from a great professional, Amar-Djalil Mezaour.

The German company AGT has as well contributed strongly, namely, from researchers Herman Ravkin, Martin Strohbach, and Panayotis Kikiras.

The Anglo-Austrian team, consisting of John Domingue, Dieter Fensel, Nelia Lasierra, Sabrina Neurerer, Serge Tymaniuk, Iker Larizgoitia, Michael Rogger, Ioannis Stavrakantonakis, Amy Strub, and Ioan Toma from STI International and University of Innsbruck, also performed really memorable work in the project.

Thanks as well to Helen Lippell from Press Association; to Axel Ngonga, Sebastian Hellman, and Sandra Praetor from the University of Leipzig; as well as to Daniel Dietrich, Pablo Mendes, and Walter Palmethofer from Open Knowledge Foundation.

Others within the National University of Ireland Galway who helped immensely are Andre Freitas, Umair ul Hassan, Aftab Iqbal, Nur Aini Rakhmawati, and Stefan Decker. The BIG project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 318062. The work has been supported in part by Science Foundation Ireland under grant SFI/12/RC/2289.

Finally, the Technical Working Groups and Sectorial Forums gratefully acknowledge the debt we owe to all our interviewees and subject matter experts who gave up their valuable time and provided extremely useful input for us. In particular, we would like to thank the following people for participating in public interviews: Matt Asay, Kevin Ashley, Sören Auer, Ricardo Baeza-Yates, François Bancilhon, Richard Benjamins, Helen Berman, Lukas Biewald, James Cheney, Hjalmar Gislason, Carole Goble, Paul Groth, Alon Halvey, Usman Haque, Steve Harris, Jim Hendler, Alek Kotcz, Prasanna Lal Das, Helen Lippell, Nick Lynch, Peter Mika, Andy Palmer, Andreas Ribbrock, Joe Sewash, John Sheridan, Kelly Stirman, Jeni Tennison, Bill Thompson, Andraž Tori, Frank van Harmelen, Marco Viceconti, Jim Webber, and Antony Williams. We would also like to thank the many CEOs, CTOs, VPs, Directors, Founders, Professors, Researchers, and Technical Experts who provided inputs to the project from organizations including AGT International, Associated Press, Atos, BBC, CGIA, Chemspider/Royal Society of Chemistry, Cosm (Pachube), CrowdFlower, Data Publica, Data Tamer, DataMarket Inc., Digital Curation Centre, European Commission, Exalead, Experian, Fraunhofer, Garlik, Geomapp, German Research Centre for Artificial Intelligence, Google, Insight, Legislation.gov.uk, MapR, MongoDB, National University of Ireland Galway, Neo, Open Data Institute, Open Knowledge Foundation, Pistoia Alliance, Press Association, Rensselaer Polytechnic Institute, Rutgers University (Protein Data Bank), SAP, Siemens, Software AG, STI International, Telefoncica, Teradata GmbH, Twitter, University of Amsterdam, University of Edinburgh, University of Innsbruck, University of Leipzig, University of Manchester, University of Sheffield, VPH Institute, VU University Amsterdam, World Bank, Yahoo!, and Zemanta. Without your support this work would not have been possible.

Ciudad Real, Spain
October 2015

José María Cavanillas

Contents

Part I The Big Data Opportunity

1	The Big Data Value Opportunity	3
	José María Cavanillas, Edward Curry, and Wolfgang Wahlster	
2	The BIG Project	13
	Edward Curry, Tilman Becker, Ricard Munné, Nuria De Lama, and Sonja Zillner	

Part II The Big Data Value Chain: Enabling and Value Creating Technologies

3	The Big Data Value Chain: Definitions, Concepts, and Theoretical Approaches	29
	Edward Curry	
4	Big Data Acquisition	39
	Klaus Lyko, Marcus Nitzschke, and Axel-Cyrille Ngonga Ngomo	
5	Big Data Analysis	63
	John Domingue, Nelia Lasierra, Anna Fensel, Tim van Kasteren, Martin Strohbach, and Andreas Thalhammer	
6	Big Data Curation	87
	André Freitas and Edward Curry	
7	Big Data Storage	119
	Martin Strohbach, Jörg Daubert, Herman Ravkin, and Mario Lischka	
8	Big Data Usage	143
	Tilman Becker	

Part III Usage and Exploitation of Big Data

9 Big Data-Driven Innovation in Industrial Sectors 169
 Sonja Zillner, Tilman Becker, Ricard Munné, Kazim Hussain, Sebnem Rusitschka, Helen Lippell, Edward Curry, and Adegboyega Ojo

10 Big Data in the Health Sector 179
 Sonja Zillner and Sabrina Neururer

11 Big Data in the Public Sector 195
 Ricard Munné

12 Big Data in the Finance and Insurance Sectors 209
 Kazim Hussain and Elsa Prieto

13 Big Data in the Energy and Transport Sectors 225
 Sebnem Rusitschka and Edward Curry

14 Big Data in the Media and Entertainment Sectors 245
 Helen Lippell

Part IV A Roadmap for Big Data Research

15 Cross-sectorial Requirements Analysis for Big Data Research 263
 Tilman Becker, Edward Curry, Anja Jentzsch, and Walter Palmetshofer

16 New Horizons for a Data-Driven Economy: Roadmaps and Action Plans for Technology, Businesses, Policy, and Society 277
 Tilman Becker, Edward Curry, Anja Jentzsch, and Walter Palmetshofer

Index 293

List of Contributors

Tilman Becker German Research Centre for Artificial Intelligence (DFKI), Saarbrücken, Germany

José María Cavanillas Atos Spain, S.A., Madrid, Spain

Edward Curry Insight Centre for Data Analytics, National University of Ireland Galway, Lower Dangan, Galway, Ireland

Jörg Daubert AGT International, Darmstadt, Germany

Nuria De Lama Atos Spain, S.A., Madrid, Spain

John Domingue STI International, Vienna, Austria; and Knowledge Media Institute, The Open University, Walton Hall, Milton Keynes, UK

Anna Fensel University of Innsbruck, Innsbruck, Austria

André Freitas Insight Centre for Data Analytics, National University of Ireland Galway, Lower Dangan, Galway, Ireland

Kazim Hussain Atos Spain, S.A., Madrid, Spain

Anja Jentsch Open Knowledge Foundation (OKF), Berlin, Germany

Nelia Lasierra University of Innsbruck, Innsbruck, Austria

Helen Lippell Press Association, London, UK

Mario Lischka acentrix GmbH, Munich, Germany

Klaus Lyko University of Leipzig, Leipzig, Germany

Ricard Munné Atos Spain, S.A., Barcelona, Spain

Sabrina Neururer Department of Medical Statistics, Informatics and Health Economics, Innsbruck Medical University, Innsbruck, Austria; and Semantic Technology Institute Innsbruck, University of Innsbruck, Innsbruck, Austria

Marcus Nitzschke University of Leipzig, Leipzig, Germany

Axel-Cyrille Ngonga Ngomo University of Leipzig, Leipzig, Germany

Adegboyega Ojo Insight Centre for Data Analytics, National University of Ireland Galway, Lower Dangan, Galway, Ireland

Walter Palmethofer Open Knowledge Foundation (OKF), Berlin, Germany

Elsa Prieto Atos Spain, S.A., Madrid, Spain

Herman Ravkin Department of Industrial Engineering, Tel-Aviv University, Ramat-Aviv, Tel-Aviv, Israel

Sebnem Rusitschka Corporate Technology, Siemens AG, Munich, Germany

Martin Strohbach AGT International, Darmstadt, Germany

Andreas Thalhammer Institute for Applied Informatics and Formal Description Methods, Karlsruhe Institute of Technology, Karlsruhe, Germany

Tim van Kasteren AGT International, Darmstadt, Germany

Wolfgang Wahlster German Research Centre for Artificial Intelligence (DFKI), Saarbrücken, Germany

Sonja Zillner Corporate Technology, Siemens AG, Munich, Germany; and School of International Business and Entrepreneurship, Steinbeis University, Berlin, Germany