DATA SCIENCE SERIES

DATA SCIENCE AND ANALYTICS STRATEGY

AN EMERGENT DESIGN APPROACH

KAILASH AWATI
ALEXANDER SCRIVEN

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

"Not only does the discipline of data science need this book, it holds critical insights and lessons for other facets of enterprise IT too. For the first time, the critically important ideas of emergent design practice have been weaved into the hyper-rational world of data science in an accessible and practical way. Kailash Awati and Alex Scriven have written the first data science book of its kind – a must read for anyone interested in the governance of data and the complex problems that data and analytics seeks to help solve."

Paul Culmsee, *Managing Partner, Seven Sigma Business Solutions*

"A refreshingly practical approach to success in data science and machine learning. The value of Awati and Scriven's contribution to this field is that emergent design lends to data science a coherence that previously was missed in the chasm between the promise of new tech and the organisational change required to harness it. They've bridged that gap with a highly accessible read, weaving the wealth of their collective experience with the rigour of leading researchers, intellectuals and practitioners into a lively jaunt covering the full vocabulary of concepts for leaders (from deep learning to tech stack to GDPR) that will hold 'aha' moments for even the most seasoned data and analytics professionals and (hopefully!) spawn a new generation of strategic leadership and emergent practice in this space."

Passiona Cottee, *Associate Director, NSW Government*

"If you are passionate about the successful implementation of data science and analytics strategies, then put this book on your required reading list. You will learn why and how to define a direction by finding and framing problems that matter to people across the organisation."

Zanne Van Wyk, *Worldwide Education Industry Architect at Microsoft*

"*Data Science and Analytics Strategy* covers a wide range of topics like building analytics and data science capability, building data driven culture in the organization and ethical aspects of practicing data science. It includes advice which are easy and very practical to use in real world scenarios. All in all, a great read for all those who want to setup analytics and data science practises within their organization."

Duhita Khadepau, *Director (Analytics and Data Science), Assignar*

"Succeeding with data science and analytics is no easy ride, however this book gives the reader a range of ideas and actions to combat the challenges faced by professionals in this field. Finding a path to success requires new approaches and this book provides a refreshing perspective for practitioners to consider as they strive for success."

Sandra Hogan, *Co-Founder, Amperfii*

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Data Science and Analytics Strategy

This book describes how to establish data science and analytics capabilities in organisations using an evolutionary approach that increases the chances of successful outcomes while minimising upfront investment. Based on their experiences and those of a number of data leaders, the authors provide actionable advice on data technologies, processes, and governance structures so that readers can make choices that are appropriate to their organisational contexts and requirements.

The book blends academic research on organisational change and data science processes with real-world stories from experienced data analytics leaders, focusing on the practical aspects of setting up a data capability. In addition to a detailed coverage of capability, culture, and technology choices, a unique feature of the book is its treatment of emerging issues such as data ethics and algorithmic fairness.

Data Science and Analytics Strategy: An Emergent Design Approach has been written for professionals who are looking to build data science and analytics capabilities within their organisations as well as those who wish to expand their knowledge and advance their careers in the data space. Providing deep insights into the intersection between data science and business, this guide will help professionals understand how to help their organisations reap the benefits offered by data. Most importantly, readers will learn how to build a fit-for-purpose data science capability in a manner that avoids the most common pitfalls.

Kailash Awati is a data and sensemaking professional with a deep interest in helping organisations tackle complex problems. He is an Adjunct Fellow in Human-Centred Data Science at the UTS Connected Intelligence Centre and a Data and Insights Manager at a government agency. Over the last decade, he has established data capabilities in diverse organisations using the principles described in this book. In addition to his work in industry, he has developed and taught postgraduate courses in machine learning and decision-making under uncertainty. He is the co-author of two well-regarded books on managing socially complex problems in organisations: *The Heretic's Guide to Best Practices* and *The Heretic's Guide to Management*.

Alexander Scriven is a senior data scientist at Atlassian in Sydney, Australia, and has experience across start-ups, government, and enterprise building analytical capacities and executing on data science projects. He greatly enjoys teaching and mentoring and has built and delivered both master's-level university courses in machine learning and deep learning and highly rated courses for online platforms such as Datacamp. His research interests are in applying data science techniques to novel industry challenges. Alex greatly enjoys bridging the gap between cutting-edge technology and business applications.

CHAPMAN & HALL/CRC DATA SCIENCE SERIES

Reflecting the interdisciplinary nature of the field, this book series brings together researchers, practitioners, and instructors from statistics, computer science, machine learning, and analytics. The series will publish cutting-edge research, industry applications, and textbooks in data science.

The inclusion of concrete examples, applications, and methods is highly encouraged. The scope of the series includes titles in the areas of machine learning, pattern recognition, predictive analytics, business analytics, Big Data, visualization, programming, software, learning analytics, data wrangling, interactive graphics, and reproducible research.

Published Titles

Public Policy Analytics
Code and Context for Data Science in Government
Ken Steif

Supervised Machine Learning for Text Analysis in R
Emil Hvitfeldt and Julia Silge

Massive Graph Analytics
Edited by David Bader

Data Science
An Introduction
Tiffany-Anne Timbers, Trevor Campbell and Melissa Lee

Tree-Based Methods
A Practical Introduction with Applications in R
Brandon M. Greenwell

Urban Informatics
Using Big Data to Understand and Serve Communities
Daniel T. O'Brien

Introduction to Environmental Data Science
Jerry Douglas Davis

Cybersecurity Analytics
Rakesh M. Verma and David J. Marchette

Exploratory Model Analysis
Explore, Explain, and Examine Predictive Models
Przemyslaw Biecek and Tomasz Burzykowski

For more information about this series, please visit: www.routledge.com/Chapman--HallCRC-Data-Science-Series/book-series/CHDSS

Data Science and Analytics Strategy

An Emergent Design Approach

Kailash Awati
Alexander Scriven

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A CHAPMAN & HALL BOOK

Designed cover image: Getty Images

First edition published 2023
by CRC Press
6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742

and by CRC Press
4 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

CRC Press is an imprint of Taylor & Francis Group, LLC

ISBN: 978-1-032-19633-6 (hbk)
ISBN: 978-1-032-19632-9 (pbk)
ISBN: 978-1-003-26015-8 (ebk)

DOI: 10.1201/9781003260158

Typeset in Palatino
by Newgen Publishing UK

Kailash

For my amazing wife who steered our family through challenging times while battling a serious illness: Arati, you inspire me every day.

Alex

For Jennifer. My world.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Contents

Foreword

In my experience the ability to go from a collection of data science use cases to a full-blown capability presents a significant challenge for many organisations. Recognising that organisations have traditionally struggled with the ability to develop and execute data and analytics strategies, this book by Kailash Awati and Alex Scriven provides practical guidance for leaders who are looking to derive tangible benefits from data. Apart from covering traditional topics such as capability development, technology, and productionising data science, a novel and perhaps the most important feature of the book is its coverage of governance and ethics. Data practitioners and enthusiasts will be exposed to the key elements of data science and analytics and will learn how to apply these features to organically and sustainably grow data capabilities.

Emergent Design urges data leaders and practitioners to look beyond the technical aspects of the discipline. Organisations that fail to recognise the sociotechnical aspects of data science will struggle to develop a sustainable data capability. Whether you are setting out on your journey or refining your existing data science programmes, there is a significant benefit to be gained by adopting the approach advocated in this book. Based on evolutionary principles of change, it provides a simple and effective framework to guide you on your data journey.

In addition to challenging conventional wisdom about how to “do data”, this book captures the diverse experiences of many data and analytics leaders who discuss how they have overcome the challenges associated with harnessing the growing potential of emerging data technologies in a scalable and sustainable manner. Many organisations are rightfully focussed on ensuring the interoperability of technology and establishing efficient, optimised processes. However, as this book stresses, the most critical element of success lies in building the skills and expertise of your greatest assets: the people in your organisation.

Craig Napier
Chief Data Officer, University of Technology Sydney
August 2022

Preface

As we were close to completing the book, *Harvard Business Review* published an article entitled, *Is Data Scientist Still the Sexiest Job of the 21st Century?*[1] The article revisits a claim made a decade ago, in a similarly titled piece about the attractiveness of the profession.[2] In the recent article, the authors note that although data science is now a well-established function in the business world, setting up the function presents a number of traps for the unwary. In particular, they identify the following challenges:

- The diverse skills required to do data science in an organisational setting
- A rapidly evolving technology landscape
- Issues around managing data science projects; in particular, productionising data science models – i.e., deploying them for ongoing use in business decision-making
- Putting in place the organisational structures/processes and cultivating individual dispositions to ensure that data science is done in an ethical manner

On reviewing our nearly completed manuscript, we saw that we have spoken about each of these issues, in nearly the same order that they are discussed in the article (see the titles of Chapters 5–8). It appears that the issues we identified as pivotal are indeed the ones that organisations face when setting up a new data science function.

That said, the approach we advocate to tackle these challenges is somewhat unusual and therefore merits a prefatory explanation.

The approach proposed in this book arose from the professional experiences of two very different individuals, whose thoughts on how to "do data" in organisational settings converged via innumerable conversations over the last five years. Prior to working on this book, we collaborated on developing and teaching an introductory postgraduate data science course to diverse audiences ranging from data analysts and IT professionals to sociologists and journalists. At the same time, we led very different professional lives, working on assorted data-related roles in multinational enterprises, government, higher education, not-for-profit organisations and start-ups. The main lesson we learned from our teaching and professional experiences is that, when building data capabilities, it is necessary to first understand where people are – in terms of current knowledge, past experience, and future plans – and grow the capability from there.

To summarise our approach in a line: *data capabilities should be grown, not grafted*.

This is the central theme of Emergent Design, which we introduce in Chapter 1 and elaborate in Chapter 3. The rest of the book is about building a data science capability using this approach.

Naturally, we were keen to sense-check our thinking with others. To this end, we interviewed a number of well-established data leaders and practitioners from diverse domains, asking them about their approach to setting up and maintaining data science capabilities. You will find their quotes scattered liberally across the second half of this book. When speaking with these individuals, we found that most of them tend to favour an evolutionary approach not unlike the one we advocate in the book. To be sure, organisations need formal structures and processes in place to ensure consistency, but many of the data leaders we spoke with emphasised the need to grow these in a gradual manner, taking into account the specific context of their organisations.

It seems to us that many who are successful in building data science and analytics capabilities tacitly use an emergent design approach, or at least some elements of it. Yet, there is very little discussion about this approach in the professional and academic literature. This book is our attempt at bridging this gap.

Although primarily written for business managers and senior data professionals who are interested in establishing modern data capabilities in their organisations, we are also speaking to a wider audience ranging from data science and business students to data professionals who would like to step into management roles. Last but not least, we hope the book will appeal to curious business professionals who would like to develop a solid understanding of the various components of a modern data capability.

That said, regardless of their backgrounds and interests, we hope readers will find this book useful … and dare we say, an enjoyable read.

Kailash Awati
Alexander Scriven
Sydney, September 2022

Notes

1 https://hbr.org/2022/07/is-data-scientist-still-the-sexiest-job-of-the-21st-century
2 https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century

Acknowledgements

Kailash and Alex

We would like to thank the data leaders who were beyond generous with their time, advice, and support in the course of writing this book. They are: Craig Napier, Duhita Khadepau, Firas Hamdan, Hema Prasad, Ian Jackman, Jan Lambrechts, Josh McNeil, Kumar Parekh, Dr Kobi Leins, Matt Minor, Sandra Hogan, Sonya Zecchin, Sylvia Jastkowiak, and Zanne Van Wyk. Brief biographies of these accomplished individuals are available in the following pages. We'd also like to thank Passiona Cottee for her significant contribution to the section on ethics in Chapter 8. Last, but not least, our thanks go out to George Knott for his support through the process of writing this book.

Kailash

I first encountered the concept of emergent design about a decade ago, in a conversation with my friend and longtime collaborator, Paul Culmsee. My thanks go out to him for introducing me to the idea and for many discussions about it in the years since.

Over the last many years, I have been fortunate to have had multiple opportunities to set up data analytics capabilities in diverse environments. It is these experiences that have enabled me to road-test and refine the ideas presented herein. I'm therefore indebted to numerous colleagues and workmates who travelled the road with me. In particular, I would like to call out Anusha, David Lefeve, Hendrik Mieves, Sean Heffernan, Homan Zhao, Sree Acharath, Sam Kohli, Nivi Srinivasan, Vidyalakshmi Rao, Matthew Harwood, Anita Menon, and Neil Finlay.

A huge thank you to my managers past and present – Mario Techera, Matthew Perry, Joe Helo, Yasuhiro Nishimi, Simon Buckingham Shum, Josh McNeil, and Celia Murphy – all of whom gave me the freedom to try out new ways of working within their organisations.

I am indebted to Simon Buckingham Shum who, in 2016, gave me the opportunity to teach in the Master of Data Science and Innovation (MDSI) programme at the University of Technology Sydney. A big thank you to all of the MDSI students I have had the privilege of teaching over the last five years: I have learnt much from my interactions with you all, both in and outside the classroom. I'd like to single out Chris Mahoney who supported me as a tutor through challenging pedagogical times during the height of the Covid pandemic. A special thanks to my brilliant co-author, Alex, from whom I have learnt much about the ins and outs of doing data science in different contexts. Yes, we are unlikely collaborators … but I think it works!

Finally, and most importantly, I'd like to thank Vikram, Rohan, and Arati, the mainstays of my life, without whom none of this (or anything else) would have been possible.

Alex

When Kailash and I began discussing how to paint this book on the blank canvas in front of us, it began with sharing stories: stories of success, stories of failure, funny stories, and frustrating stories. All manner of stories from the quite different paths we have walked.

It is therefore important that I acknowledge and thank everyone for the roles they played in these stories which have shaped myself and my input into this project. Firstly, a broad thanks to everyone who has been involved in working with, for, and around me during my career. Special thanks to Sally Wade, Gelina Talbot, and Jon Beard (as well as my old managers Mick G, Jen M.W, Mel B, and others) for the belief in me and opportunities they created and supported me through. To Adrian Cordiner (and the Digital Rhinos) for all our work (past, present, and future!) and his support during my studies. As well as the entire "Datanauts" crew for the projects we worked on together.

I would also like to thank the various people I have worked with on teaching and academic work across UNSW and UTS, especially Isabella Dobrescu, Alberto Motta, and my current research team (notably Bogdan, Kaska, and David). Additionally, all the students whom I have taught including those in the Master of Data Science and Innovation (MDSI) programme at the University of Technology Sydney. I still feel ongoing pride seeing all the amazing things they all continue to do.

An important acknowledgement and thanks is needed for ongoing friend and mentor figures Perry Stephenson, Tim Cubitt, and Kailash (of course, an additional thanks for being my co-author on this project) as I've wended and wound my way through my career. A special thanks to Duhita Khadepau for her ongoing friendship, mentorship, and for everything we have worked on together, including her involvement in this project.

Most importantly, the biggest thanks to my close friends, to my family, to my constant companion Alfie, and most especially to my rock, Jennifer. This project and everything that I have done which is represented here I couldn't have done without your support and care.

Contributors

Passiona Cottee is a leader at the nexus of data, automation, privacy, and ethics, drawing on over 20 years' experience across both public and private sectors. A machine learning teacher and data ethicist, she is currently setting up an AI lab function within a foresight team at the NSW Department of Premier and Cabinet and guest lectures in the University of Technology Sydney's Data Science for Innovation program.

Firas Hamdan is currently the Data Governance Lead at KPMG Australia and has over 20 years' experience working with clients assessing governance practices and creating practical programmes to build a lasting culture of governance and ethical data practices. Firas has a long history of successfully leading data analytics and governance programmes across a variety of industries including banking and financial services, telecommunication, media, retail, energy, utilities, and assets management.

Sandra Hogan is an accomplished data science and analytics leader with over 20 years' experience in building and leading analytics and data science teams and over 10 years in providing specialist consulting to many large organisations across Australia and New Zealand. In 2018, Sandra was recognised as #2 Analytics Leader in Australia by the Institute of Analytics Professionals of Australia (IAPA). She is the founder of Leap Analytics, an analytics advisory company, and a co-founder of a software company, Amperfii, helping people to collaborate and create great outcomes with data and analytics.

Ian Jackman is currently the General Manager, Data and Analytics (CDO) at Bendigo and Adelaide Bank, where he has developed and driven the organisational data and analytics strategy, embedding data governance, and a data-driven culture whilst enabling value from data through an enterprise analytics centre of excellence. He has deep expertise in analytics and technology, and has consulted to organisations across industries to derive business and customer value.

Sylvia Jastkowiak is Principal Advisor in Data Ethics and Privacy at NEC Australia where she has worked with a variety of organisations and industries. She has specialist expertise in designing privacy tools and ethical frameworks that help organisations navigate their legal and social obligations when utilising personal and sensitive data.

Duhita Khadepau is currently the Director (Analytics and Data Science) at Assignar, a construction-tech start-up in Sydney, Australia. She is a seasoned data professional with over 15 years' experience in designing and executing data strategies of organisations in India, Singapore, and Australia, across the homeware, banking, B2B SaaS, and retail e-commerce sectors. She has held pivotal roles in building data capacities across some of Australia's most successful start-ups, helping them scale and raise funds, and has held lecturing and teaching positions in data science.

Jan Lambrechts is currently a data ethics lead at ANZ with deep prior experience across both government and commercial organisations in data ethics, privacy, data governance, and overall analytics enablement. He has led teams and projects to build ethical frameworks from the ground up and top-down, and worked closely with teams to translate and embed these into everyday practice.

Kobi Leins (GAICD) is an Honorary Senior Fellow of King's College, London; Non-Resident Fellow of the United Nations Institute for Disarmament Research; Advisory Board Member of the Carnegie AI and Equality Initiative; Member of Standards Australia as a technical expert on the International Standards Organisation's work on AI Standards; co-founder of Responsible Innovation and the Life Sciences with IEEE; and Affiliate, ARC Centre of Excellence for Automated Decision-Making and Society. Leins has previously managed programmes and teams in administrative law and justice, humanitarian law, human rights law, and disarmament with the UN and the International Committee of the Red Cross and worked in two different university faculties of Engineering and Computer Science. Leins is the author of *New War Technologies and International Law: The Legal Limits to Weaponising Nanomaterials*, Cambridge University Press (2022). Further publications can be found at kobileins.com.

Josh McNeil is the Manager (Direct Marketing and Campaigns) at Cancer Council NSW and has over a decade of experience in the NFP space in using data to drive organisational outcomes and building analytics capabilities.

Matt Minor is currently the Head of Group Data and Analytics at the Blackmores group where he has built and executed on building their analytics strategy and capacity. He has deep experience building and leading analytics teams across Australia and the UK, across industries including finance, software, and health.

Craig Napier is currently the Chief Data Officer at the University of Technology, Sydney, and is the President of the Data Warehouse Institute (Australian Chapter). In 2020, he was recognised as being amongst the top three analytics leaders by the Institute of Analytics Professionals Australia (IAPA). He has led and developed analytics strategies and teams driving enterprise data initiatives at a number of Australia's top universities and held lecturing positions in business analytics.

Kumar Parekh is an experienced leader in data and AI strategy who has built data strategies and overseen implementation of analytics work across the entire business value chain. His career has included building data capabilities in major institutions across industries such as government (state and federal), mining, financial services, higher education, aged care, airlines, and property to name a few. He was at the helm of one of the earliest pioneering data teams in geocoding data in Australia.

Hema Prasad is a seasoned data and analytics executive leader with over two decades of global experience enabling large-scale business transformations via data, analytics, and business intelligence for Fortune-50 and ASX clients in USA, India, and Australia. Specialising in data strategy, data platforms modernisation, data architecture, governance, big data, cloud and platform engineering, data science, and analytics, Hema is a highly regarded and recognised data practitioner. An active public speaker, industry advisor, and mentor, she is passionate about working with students and aspiring data professionals to raise awareness on data and analytics career pathways.

Zanne Van Wyk is a data and analytics strategist with over 30 years' experience in building and executing data analytics and data science strategies across a variety of industries across the globe, with special expertise in the education sector. She is currently a worldwide education industry architect at Microsoft, developing strategies and roadmaps for a portfolio of industry partners.

Sonya Zecchin has led and built analytics capabilities across industries as varied as insurance, telecommunications, government, and logistics over the past 25 years. She has held lecturing positions on applying analytics to the business context and has been a learning content developer for IAPA. Her focus is now on supporting the development of people in analytics by delivering original and curated learning experiences.

1

Introduction

Data Science as a Sociotechnical Capability

This book is the outcome of over 35 years of collective experience gained by us, the authors, in building data capabilities in diverse organisations, ranging from large multinationals and technology companies to fast-growing start-ups and small not-for-profits. Our key message is that building an effective data capability[1] is, at its heart, a problem of organisational change. This may come as a surprise to some readers, but we suspect it will strike a chord with those who have walked the path of building such capabilities.

As noted by many researchers (e.g. Gagné et al. 2000, Peus et al. 2009), the success of an organisational change initiative depends on managing its impact on those who are affected by it. Consequently, setting up a data capability in an organisation requires a careful consideration of people, technology, and, more importantly, the interactions between the two. One could say that a data science capability is a *sociotechnical*[2] system.

The introduction of a new way of working into an organisational setting will typically trigger reactions ranging from unbridled enthusiasm to covert resistance. Often the enthusiasts (typically managers) will not understand the reluctance of the resistors (typically employees who feel threatened or undermined). These issues must be worked through collectively as the strategy is implemented. Indeed, resolving such issues is just as important as the more obvious decisions pertaining to skills and technologies.

The Fallacy of Strategic Alignment

The conventional approach to formulating and implementing a data science strategy takes a technocentric view, i.e., it focuses on:

DOI: 10.1201/9781003260158-1

a. Acquiring the appropriate technology and skills.
b. Aligning these to the business. This is sometimes referred to as strategic alignment.

Given adequate budget, technology and skills are easy enough to acquire in the market. Strategic alignment is considerably harder, and it's worth understanding why.

The notion of strategic alignment of technology can be traced back to an influential paper by Henderson and Venkatraman (1989). The paper describes the need for "aligning" the business and technology strategies of companies and details a "Strategic Alignment Model" that purports to "guide management practice" towards achieving this. Although it is acknowledged that data science is distinct from other technology capabilities, in practice much of data science strategy work takes its inspiration from this notion of alignment. For example, as Robert de Graaf notes in his book *Managing Your Data Science Projects* (de Graaf 2019):

> *It's crucial for data science teams to be able to demonstrate their value by linking their activities to their organization's mission because there will be doubters who believe that data science is a waste of time.*

And a bit later:

> *The first step in developing a [data science] team strategy that achieves alignment with the organization is a careful study of the organization's strategy document… Every organization has slightly different overall goals and mission… The challenge for the data science team is to decide how the data science team supports those goals.*

The notion of alignment assumes that data science is a capability that can be grafted on to an organisation rather than grown. This book argues the opposite: *that data science is a capability that is best grown rather than grafted*. Growing a capability is a matter of *evolution*, not alignment. Indeed, alignment of the kind de Graaf mentions is difficult to achieve because the real world has a logic of its own that tends to escape our strategy documents and plans. So, as we will discuss later (Chapter 4), your first step is not about studying organisational strategy documents as de Graaf suggests; it is about defining a direction by finding and framing problems that matter to people across the organisation. If done right, the "alignment" that strategists hanker after will come for free. Our job in this book is to show you a way to "do it right".

The notion of strategic alignment has been problematised by many academics, but none have done so as eloquently as the late Claudio Ciborra, who wrote:

> *… while strategic alignment may be close to a truism conceptually, in the everyday business it is far from being implemented. … If alignment was supposed to be the ideal "bridge" connecting the two key variables [business and technology], it must be admitted that such a conceptual bridge faces the perils of the concrete bridge always re-designed and never built between continental Italy and Sicily, (actually, between Scylla and Charybdis) its main problem being the shores: shifting and torn by small and big earthquakes … .*
>
> *Ciborra 1997*

The problem with the notion of strategic alignment is that it is an abstraction that is not "out there in the world" but resides in the heads of strategists and the documents they produce. Plans that attempt to realise this kind of alignment fall apart as soon as they encounter messy real-world details. This tension between plans and reality is one of the many paradoxes inherent in strategy development. We will revisit this in Chapter 3 where we describe the principles of Emergent Design. The main point we wish to make now is that it is best to avoid making detailed plans too early in the game. We realise that this may go against everything you've heard or read about strategy development, so here are a couple of stories that might help illustrate our point. Along the way, we'll also take the opportunity to introduce a few concepts that are central to the approach we advocate in this book.

A Tale of Two Databases[3]

Many years ago, Kailash joined an organisation in the throes of change. Among other things, they were in the process of replacing a venerable Lotus Notes-based system with a newer Customer Relationship Management (CRM) product. As a part of this effort, he was asked to build a system to integrate data from the CRM system with other syndicated and publicly available datasets. The requirements were complex but, fortunately, the development team and key business stakeholders were co-located on one floor. The system design evolved through continual, often animated, discussions over the entire development period, in an environment characterised by openness and trust.

The system was delivered on schedule, with minimal rework required.

Some years later, he was invited to participate in a regional project aimed at building a data warehouse[4] for subsidiaries across Asia. The initiative was driven by the corporate IT office located in Europe. This was part of a new organisation-wide strategy to harmonise a data landscape that was – to put it mildly – messy. However, the subsidiaries thought their local systems were just fine. They were suspicious of corporate motives which they saw as a power play that would result in loss of autonomy over data and reporting.

After much debate in many videoconferences, a face-to-face meeting was called to resolve the issue.

The Wickedness of Building Data Capabilities

A few weeks before the meeting, Kailash stumbled on a 1973 paper by Horst Rittel, a professor of design at UC Berkeley (Rittel and Webber 1973). In the paper, Rittel coined the phrase *wicked problem* to describe a complex situation that is perceived in different ways by different stakeholders and is therefore difficult to translate into a clear problem statement. Rittel described ten characteristics of such problems. Although Rittel was talking about problems relating to town and infrastructure planning, Kailash saw that the characteristics he described applied exactly to the problem of the data warehouse. The first column in Table 1.1 lists the characteristics according to Rittel and the second describes the relevance of the characteristic to Kailash's data warehousing dilemma.

A question: how many of the characteristics in Table 1.1 apply to complex data projects you have worked on?

When dealing with wicked problems, the trick is to find a way to surface and reconcile diverse viewpoints. One therefore needs to make multiple perspectives explicit in a manner that enables a group to develop a shared understanding of contentious issues. Done right, this can lead to a resolution of the issue, at least partially. In other words, it is about seeking diverse viewpoints on the issue with a view to finding common ground, however small. Our point in telling such a story at this early stage is to highlight the wicked aspects of building new data capabilities in organisations.

As the story about the second database illustrates, *data is invariably political*. Often, the department or function that collects or generates the data will, by default, be seen as the "owner" of the data. That is, they have the mandate to tell the story of that data and use it for reporting, forecasting, or modelling. More importantly, they determine who gets access to the data. It is important for strategists and analysts to be sensitive to potential political issues such as ownership. Data professionals who overlook this aspect of their work will come unstuck for reasons that have nothing to do with technical competence.

There is another, less obvious, political issue that is worth unpacking. It has to do with how the data is interpreted and whose interests the interpretation represents. The mainstream approach to data modelling assumes that real-world objects and relationships can be accurately represented by models. As an example, a data model representing a sales process might consist of entities such as customers and products and their relationships, such as sales (customer X purchases product Y). It is tacitly assumed that objective,

TABLE 1.1

Characteristics of Wicked Problems and Their Relevance to Data Warehousing

Wicked Problem Characteristic	Relevance to the Situation
There is no definitive formulation of a wicked problem.	The formulation of the problem – standardisation vs reporting – depended on who was asked.
Wicked problems have no stopping rule.	A data warehouse is never done; it evolves as user requirements evolve.
Solutions to wicked problems are not true-or-false, but good-or-bad.	Data architecture is an exercise in compromise – there is no absolute right or wrong.
There is no immediate or ultimate test of a solution to a wicked problem.	Since a perfect solution does not exist, there is no ultimate test of a solution.
Every solution to a wicked problem is a "one-shot operation" because there is no opportunity to learn by trial and error; every attempt counts significantly.	Though this is not always true, there are invariably some data design decisions that can be extremely expensive or even impossible to fix without redoing the entire thing.
Wicked problems do not have an enumerable (or an exhaustively describable) set of potential solutions, nor is there a well-described set of permissible operations that may be incorporated into the plan.	There are, in principle, a huge number of viable data warehouse designs.
Every wicked problem can be considered to be a symptom of another problem.	There are fundamental principles of data warehouse design, but each data warehouse is unique, reflecting the unique requirements and technology choices of the organisation.
The existence of a discrepancy representing a wicked problem can be explained in numerous ways. The choice of explanation determines the nature of the problem's resolution.	The discrepancy here was that we had two conflicting approaches to designing the data warehouse. However, this was a political issue, not a technical one, and the politics itself was due to differences in perception of the situation.
The planner has no right to be wrong.	The database designer would be held responsible for the consequences of design decisions that were made. The decision itself was about steering a narrow course between the Scylla of subsidiaries and the Charybdis of corporate.

bias-free models of entities and relationships of interest can be built by asking the right questions and using appropriate information collection techniques.

However, things are not quite so straightforward: as data professionals know, real-world data models are invariably tainted by compromises between rigour and reality. This is inevitable because the process of building a data model involves at least two different sets of stakeholders whose interests are often at odds – business users and data modelling professionals. The former are not interested in the purity of model; they care about how well it supports their business processes. The interests of the latter, however, are often the opposite. And if that wasn't enough, there is the interest of the customer as well – for example, about how their data is protected from unauthorised access and the potential for misuse. We'll discuss these issues at length in Chapter 8.

The above reveals a truth about data modelling that is not fully appreciated by practitioners: that it is a process of negotiation and ethics rather than a search for a true representation of business reality. In other words, it is a sociotechnical problem that has wicked elements. This point has been highlighted in a brilliant paper by Heinz Klein and Kalle Lyytinen (1992). The key takeaway from the paper is that a data model is but one possible interpretation of reality. As such, there are many possible interpretations of reality so the "correctness" of any model hinges not on some objective truth but on a negotiated, best-for-group interpretation. This necessarily implies that a well-constructed data model "fuses" or "brings together" at least two different interpretations – those of users and modellers.

The mainstream view of data is that it asserts a truth and that data models reflect that truth. The view we are describing here, however, makes us aware that *data models are built in such a way as to support particular agendas*. Moreover, since the people who use the model are not those who construct it, *a gap between assumed and actual meaning is inevitable*. Indeed, even meaning evolves over time as the design evolves. It has been noted that a good design not only implements current business processes but also facilitates change (Dorst 2019). This necessarily implies that the design itself should be capable of evolving as the group's understanding of context evolves.

It is worth pausing here to think about the implications of the above. The example we have discussed deals with a technical matter – the design of a data warehouse. Even so, we see that there are wider issues that need to be addressed before the technologists can get to work. The point we want to emphasise here is that *social and political considerations permeate the entire spectrum of data capability building, from the technical to the organisational*.

The Notion of Emergent Design

Some years after the data warehousing story related above, Kailash found himself in a more senior role in the same organisation, a role in which he had responsibility for data-related work across a geographic region.

When Kailash was promoted to the new role, his boss asked him to explore the possibility of setting up a regional development centre for analytics that could serve the entire organisation. There was a clear cost argument in favour of such centre. However, given the sharply divided opinions around offshoring, the public airing of such a proposal would cause all kinds of reactions, many of which would be negative. The first problem was to address those upfront.

Kailash talked to the usual suspects, a few big outsourcers and consultancies, but soon realised that their aims were not congruent with his. Everything the outsourcers said pointed to high costs and potential conflicts down the line, such as vendor lock in and expensive contract variations.[5] As we will elaborate later in Chapter 6, the hidden costs of outsourcing are much too high. This was the second problem.

Oh, and if that weren't enough, there was another catch: the boss told him that there was zero budget for this at the time as it was not an official project. Moreover, since the initiative did not exactly have backing from corporate, it was unlikely to be supported in the near future. This was problem number three.

It was around that time that Kailash came across the notion of *Emergent Design* (Cavallo 2000a, 2000b), the key theme behind this book. The essential idea is to start from where people are and take small steps, each of which leads to demonstrable improvement. This generally requires some trial and error, but since the investment at each step is small and the benefit is demonstrable, it is not hard to convince the folks who sign cheques. Moreover, this enables one to continually adjust one's approach based on feedback from the previous step, much like nature does in the process of evolution. Actions are based on a given context, but the context itself changes because of the action and thus necessitates recalibrating subsequent actions.

At this point you might be thinking "this is exactly the same as an Agile approach – such as Scrum". That is not so. The key difference between Agile and Emergent Design is that in Agile, the endpoint of a sprint (or whatever else it is called in your favourite flavour of Agile) is well defined; in Emergent Design, it isn't. In the latter, we set out in a particular direction but without a well-defined endpoint. Indeed, there is no endpoint because new horizons open up as one proceeds, leading to new goals. This is exactly what one wants from a good strategy; a strategy that cannot evolve is worthless.

We'll say more about Emergent Design and how it was used to address the above challenge in Chapter 3. For now, we will make the observation that the way a problem is *framed*[6] – be it building a data warehouse or a data capability or anything else – is based on a range of implicit assumptions about the underlying nature of the problem. Most often the assumptions are based on conventional wisdom or "best practice" thinking, which leads to canned solutions that are rarely suited to the context at hand. Indeed, every organisation is unique in its details, so canned or one-size-fits-all approaches are unlikely to work well. Finding the right solution is a matter of taking small steps, each of which makes a tangible difference to the business. This is a process of *wayfinding* – setting out in a direction and working one's way to a destination, with details of next steps becoming apparent only as one progresses. This is not to say that one is proceeding blind. The metaphor we like to use is that it is akin to finding one's way through a thick fog; you need to focus on the immediate next step because your visibility is limited. Conventional strategies assume that one has a clear view of the future. In reality, the future is always foggy, which is why we advocate Emergent Design.

Given that data is political and developing a strategy is a process of wayfinding, one is inexorably led to the conclusion that *building new data capabilities in an organisation is an emergent process*. A technocentric approach that focuses largely on technical aspects such as technical knowledge, standards, and infrastructures exclusively will lead to disappointment. This is not to say that these are unimportant. Rather it is that they must be decided based on a deep understanding of the wider organisational context in which they will be implemented.

What to Expect from This Book

Given the above introduction, you're probably wondering what *actionable* advice you are going to get from this book. It should be clear that we are not going to provide you with a formula complete with templates and roadmaps based on "best practices" that you can copy-paste to your situation. Indeed, our intention is the opposite: to emphasise that there are no best practices … but there are good practices. So, what is a good practice?

In the context of building a data capability, good practice lies in:

1. Understanding the current state of your organisation from the perspective of how data is currently used and how it could be used.
2. Using the understanding developed in (1), to formulate high-level aspirational goals that set a direction rather than an objective. It is trivially true that you cannot foresee the future in all its detail. If this is so, then you cannot know upfront where your organisation is going to end up, so it is pointless to try and articulate that objective at the start. *Focus on the journey instead.*
3. What does "focusing on the journey" entail? Essentially it means eschewing big changes in favour of incremental and adaptive improvements.
4. Above all, putting people at the heart of what you do.

The value of a strategy is not in the strategy itself but in the process of strategising – thinking about where you are right now and (keeping that in mind) what should your next move be. By the end of this book, we hope to convince you that not only is this a practical approach, but is one that is superior to the conventional approach to developing data or any other sociotechnical capabilities.

The Structure of the Book

Since the approach we propose is novel, we will shift between theory and practice. As you go through the book, you will notice the chapters will (sort of) alternate between the two. That said, even the theoretical chapters are grounded via real-world examples and case studies. We hope these will help clarify what Emergent Design is and why we think it is the best way to go about building sociotechnical capabilities. Here's a brief summary of the structure of the book.

Chapter 2 is an introduction to data science for managers. Chapter 3 provides readers with a detailed introduction to Emergent Design. Chapter 4 describes the first – and most important – step in formulating a data science strategy based on the principles of Emergent Design. Chapters 5–8 cover in detail various aspects of the strategy including capability and culture (Chapter 5), technical matters (Chapter 6), an end-to-end view of the data science workflow (Chapter 7), data & AI governance, ethics, and privacy (Chapter 8), and finally a closing chapter to summarise the key points and offer some tips on selling the approach to your executives (Chapter 9).

Throughout we provide a number of vignettes based on our experiences of building data capabilities in diverse organisations and, more importantly, those of accomplished data leaders whose biographies appear in the front matter. It should be noted that not all the elements discussed in this book will be relevant for your situation. Feel free to pick and choose the bits you think will be useful in your context. Akin to evolution, Emergent Design is ultimately about doing what helps you make progress, however small.

We hope the stories related in this chapter and the commentary around them illustrate the limitations of the conventional, technocentric approach to building data capabilities in organisations. Along the way, we have also taken the opportunity to introduce a couple of concepts that are key threads that run through the book: *wicked problems* and *Emergent Design*. Our aim was to give you a sense of the approach advocated in this book. If you're browsing this chapter in your local bookstore, library, or online, we hope what you have read so far has piqued your interest enough to take this book home and read further.

Notes

1 Note that we will use the phrases *data capability* and *data science capability* interchangeably in this book as we will cover both. However, our primary focus is on building a *modern* data capability, which is necessarily about data science and analytics.
2 The term "sociotechnical system" came from the work done by Eric Trist and Ken Bamforth in the 1940s and 1950s, on the interactions between workers and technology in coal mines in the UK (Trist and Bamforth 1951). A brief account of the early history of the term can be found at: https://eight2late.wordpress.com/2015/04/07/from-the-coalface-an-essay-on-the-early-history-of-sociotechnical-systems/. In recent years, the term has been picked up and used in myriad other contexts (see Jasanoff and Kim 2013, for example).
3 This section and the following one are adapted from Awati (2021).
4 We will explain this term in Chapter 2. For now, think of it as a database that integrates data from several different systems.
5 https://eight2late.wordpress.com/2016/05/03/the-hidden-costs-of-it-outsourcing/
6 Problem framing is about defining or extracting a problem from a business situation. We'll say more about this at various points in this book.

References

Awati, K. (2021), "Software development as a wicked problem", *Increment*, Issue 19. Available online at: https://increment.com/planning/software-development-as-a-wicked-problem/

Cavallo, D. (2000a), "Emergent design and learning environments: Building on indigenous knowledge", *IBM Systems Journal*, Vol. 39 No. 3–4, pp. 768–781.

Cavallo, D. (2000b), *Technological Fluency and the Art of Motorcycle Maintenance: Emergent Design of Learning Environments* (Doctoral dissertation, Massachusetts Institute of Technology). Available online at: https://dspace.mit.edu/bitstream/handle/1721.1/9135/45233377-MIT.pdf

Ciborra, C. (1997), "De profundis? Deconstructing the concept of strategic alignment", *Scandinavian Journal of Information Systems*, Vol. 9 No. 1, pp. 67–82.

de Graaf, R. (2019), *Managing Your Data Science Projects*. Apress, Berkeley, CA.

Dorst, K. (2019), "Design beyond design", *She Ji: The Journal of Design, Economics, and Innovation*, Vol. 5 No. 2, pp. 117–127.

Gagné, M., Koestner, R. and Zuckerman, M. (2000), "Facilitating acceptance of organizational change: The importance of self-determination", *Journal of Applied Social Psychology*, Vol. 30 No. 9, pp. 1843–1852.

Henderson, J. and Venkatraman, N. (1989), "Strategic alignment: A framework for strategic information technology management", MIT Centre for Information Systems Research, Working Paper No. 190. Available online at: https://dspace.mit.edu/bitstream/handle/1721.1/49117/strategicalignme00hend.pdf

Jasanoff, S. and Kim, S. H (2013), "Sociotechnical imaginaries and national energy policies", *Science as Culture*, Vol. 22 No. 2, pp. 189–196.

Klein, H. K. and Lyytinen, K. (1992), "Towards a new understanding of data modeling", in Floyd, C., Zuellighoven, H., Budde, R. and Keil-Slawik, R. (Eds), *Software Development and Reality Construction*, Springer Verlag, Berlin, pp. 203–219.

Peus, C., Frey, D., Gerkhardt, M., Fischer, P. and Traut-Mattausch, E. (2009), "Leading and managing organizational change initiatives", *Management Revue*, Vol. 20 No. 2, pp. 158–175.

Rittel, H. W. and Webber, M. M. (1973), "Dilemmas in a general theory of planning", *Policy Sciences*, Vol. 4 No. 2, pp. 155–169.

Trist, E. and Bamforth, K. (1951), "Some social and psychological consequences of the longwall method of coal-getting: An examination of the psychological situation and defences of a work group in relation to the social structure and technological content of the work system", *Human Relations*, Vol. 4 No. 2, pp. 3–38.

Introduction

Awati, K . (2021), “Software development as a wicked problem”, Increment, Issue 19. Available online at: https://increment.com/planning/software-development-as-a-wicked-problem/

Cavallo, D . (2000a), “Emergent design and learning environments: Building on indigenous knowledge”, IBM Systems Journal, Vol. 39 No. 3–4, pp. 768–781.

Cavallo, D . (2000b), *Technological Fluency and the Art of Motorcycle Maintenance: Emergent Design of Learning Environments* (Doctoral dissertation, Massachusetts Institute of Technology). Available online at: https://dspace.mit.edu/bitstream/handle/1721.1/9135/45233377-MIT.pdf

Ciborra, C . (1997), “De profundis? Deconstructing the concept of strategic alignment”, Scandinavian Journal of Information Systems, Vol. 9 No. 1, pp. 67–82.

de Graaf, R . (2019), Managing Your Data Science Projects. Apress, Berkeley, CA.

Dorst, K . (2019), “Design beyond design”, She Ji: The Journal of Design, Economics, and Innovation, Vol. 5 No. 2, pp. 117–127.

Gagné, M ., Koestner, R . and Zuckerman, M . (2000), “Facilitating acceptance of organizational change: The importance of self-determination”, Journal of Applied Social Psychology, Vol. 30 No. 9, pp. 1843–1852.

Henderson, J . and Venkatraman, N . (1989), “Strategic alignment: A framework for strategic information technology management”, MIT Centre for Information Systems Research, Working Paper No. 190. Available online at: https://dspace.mit.edu/bitstream/handle/1721.1/49117/strategicalignme00hend.pdf

Jasanoff, S . and Kim, S. H. (2013), “Sociotechnical imaginaries and national energy policies”, Science as Culture, Vol. 22 No. 2, pp. 189–196.

Klein, H. K. and Lyytinen, K . (1992), “Towards a new understanding of data modeling”, in Floyd, C ., Zuellighoven, H ., Budde, R. and Keil-Slawik, R . (Eds), Software Development and Reality Construction , Springer Verlag, Berlin, pp. 203–219.

Peus, C ., Frey, D ., Gerkhardt, M ., Fischer, P . and Traut-Mattausch, E . (2009), “Leading and managing organizational change initiatives”, Management Revue, Vol. 20 No. 2, pp. 158–175.

Rittel, H. W. and Webber, M. M. (1973), “Dilemmas in a general theory of planning”, Policy Sciences, Vol. 4 No. 2, pp. 155–169.

Trist, E . and Bamforth, K . (1951), “Some social and psychological consequences of the longwall method of coal-getting: An examination of the psychological situation and defences of a work group in relation to the social structure and technological content of the work system”, Human Relations, Vol. 4 No. 2, pp. 3–38.

What Is Data Science?

Dama International (2017), DAMA-DMBOK: Data Management Body of Knowledge. (2nd ed.). Technics Publications, LLC, Denville, NJ.

Kimball, R . and Ross, M . (2011), The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling. John Wiley & Sons, Indianapolis, IN.

Shearer, C . (2000), “The CRISP-DM model: The new blueprint for data mining”, Journal of Data Warehousing, Vol. 5 No. 4, pp. 13–22.

The Principles of Emergent Design

Allison, I . and Merali, Y . (2007), “Software process improvement as emergent change: A structurational analysis”, Information and Software Technology, Vol. 49 No. 6, pp. 668–681.

Baptista, J ., Stein, M. K. , Klein, S ., Watson-Manheim, M. B. and Lee, J . (2020), “Digital work and organisational transformation: Emergent digital/human work configurations in modern organisations”, The Journal of Strategic Information Systems, Vol. 29 No. 22, pp. 1–10.

Barabba, V. , Pourdehnad, J . and Ackoff, R. L. (2002), “On misdirecting management”, Strategy and Leadership, Vol. 30 No. 5, pp. 5–9.

Bateson, G . (1972), Steps to an Ecology of Mind, University of Chicago Press, Chicago, IL.
Bateson, G . (1979), Mind and Nature: A Necessary Unit, Dutton, New York, NY, p. 4.
Cavallo, D . (2000a), “Emergent design and learning environments: Building on indigenous knowledge”, IBM Systems Journal, Vol. 39 No. 3–4, pp. 768–781.
Cavallo, D . (2000b), Technological Fluency and the Art of Motorcycle Maintenance: Emergent Design of Learning Environments (Doctoral dissertation, Massachusetts Institute of Technology). Available online at:
https://dspace.mit.edu/bitstream/handle/1721.1/9135/45233377-MIT.pdf
Chia, R. C. and Holt, R . (2009), Strategy without Design: The Silent Efficacy of Indirect Action, Cambridge University Press, Cambridge, UK.
Culmsee, P . and Awati, K . (2013), The Heretic’s Guide to Best Practices: The Reality of Managing Complex Problems in Organisations, iUniverse Star, Indianapolis, IN.
Culmsee, P . and Awati, K . (2016), The Heretic’s Guide to Management: The Art of Harnessing Ambiguity, Heretics Guide Press, Marsfield, Australia.
Dallas, L. L. (2011), “Short-termism, the financial crisis, and corporate governance”, The Journal of Corporation Law Vol. 37, p. 267.
Davies, R ., Haldane, A. G. , Nielsen, M . and Pezzini, S . (2014), “Measuring the costs of short-termism”, Journal of Financial Stability, Vol 12, pp. 16–25.
Edmondson, A . (1999), “Psychological safety and learning behavior in work teams”, Administrative Science Quarterly, Vol. 44 No. 2, pp. 350–383.
Kauffman, S . (2014), “Prolegomenon to patterns in evolution”, Biosystems, Vol. 123, pp. 3–8.
Kauffman, S . and Garre, A . (2015), “Beyond Descartes and Newton: Recovering life and humanity”, Progress in Biophysics and Molecular Biology, Vol. 119 No. 3, pp. 219–244.
Kay, J . (2012), Obliquity: Why Our Goals Are Best Achieved Indirectly, Penguin Books, Reprint Edition, London, UK.
Klein, G . (2007), “Performing a project premortem”, Harvard Business Review, Vol. 85 No. 9, pp. 18–19.
Kurtz, C. F. and Snowden, D. J. (2003), “The new dynamics of strategy: Sense-making in a complex and complicated world”, IBM Systems Journal, Vol. 42 No. 3, pp. 462–483.
Laverty, K ., (2004), “Managerial myopia or systemic short-termism? The importance of managerial systems in valuing the long term”, Management Decision, Vol. 42 No. 8, pp. 949–962.
Luna-Reyes, L. F. , Zhang, J ., Ramon Gil-Garcia, J . and Cresswell, A. M. (2005), “Information systems development as emergent socio-technical change: A practice approach”, European Journal of Information Systems, Vol. 14 No. 1, pp. 93–105.
Maltezos, S . and Georgakopoulou, A . (2021), “Novel approach for Monte Carlo simulation of the new COVID-19 spread dynamics”, Infection, Genetics and Evolution, Vol. 92, p. 104896.
Nadler, G . and Hibino, S . (1990), Breakthrough Thinking, Prima Publications and Communications, Indianapolis, IN.
Reason, J . (2000), “Human error: Models and management”, British Medical Journal, Vol. 320 No. 7237, pp. 768–770.
Rittel, H. W. J. (1972), “On the planning crisis: Systems analysis of the first and second generations”, Bedriftsokonomen, Vol. 8, pp. 390–396.
von Foerster, Heinz (2003), “Ethics and second-order cybernetics”, in Understanding Understanding. Springer, New York, NY, pp. 287–304.
Whitehead, A. N. (1925), Science and the Modern World: Lovell Lectures, Macmillan, New York, NY.
Williamson, O. E. (2010), “Transaction cost economics: The natural progression”, American Economic Review, Vol. 100 No. 3, pp. 673–690.
Winograd, T . and Flores, F ., (1986), Understanding Computers and Cognition: A New Foundation for Design, Ablex Publishing, Norwood, NJ.
Xie, G . (2020), “A novel Monte Carlo simulation procedure for modelling COVID-19 spread over time”, Scientific Reports, Vol. 10 No. 1, pp. 1–9.

Charting a Course

Birkinshaw, J . and Ridderstråle, J . (1999), “Fighting the corporate immune system: A process study of subsidiary initiatives in multinational corporations”, International Business Review, *8* (2), pp. 149–180.

Culmsee, P . and Awati, K . (2014), “The map and the territory: A practitioner perspective on knowledge cartography”, in Okada, A ., Buckingham Shum, S . and Sherborne, T . (Eds), Knowledge Cartography: Advanced Information and Knowledge Processing, Springer, London, pp. 261–292. doi: 10.1007/978-1-4471-6470-8_12

Culmsee, P . and Awati, K . (2016), The Heretic’s Guide to Management: The Art of Harnessing Ambiguity, Heretics Guide Press, Marsfield, Australia.

Nadler, G . and Hibino, S . (1990), Breakthrough Thinking, Prima Publications and Communications, Indianapolis, IN.

Capability and Culture

Culmsee, P . and Awati, K . (2016), The Heretic’s Guide to Management: The Art of Harnessing Ambiguity, Heretics Guide Press, Marsfield, Australia.

Goldratt, E . (1990), The Haystack Syndrome: Sifting Information Out of the Data Ocean, North River Press, Great Barrington, MA.

Hackman, J . (2012), “From causes to conditions in group research”, Journal of Organizational Behavior, Vol. 33 No. 3, pp. 428–444. https://doi.org/10.1002/job.1774

Van Gelder, T . (2005), “Teaching critical thinking: Some lessons from cognitive science”, College Teaching, Vol. 53 No. 1, pp. 41–48.

Technical Choices

Awati, K . and Nikolova, N . (2022), “From ambiguity to action: integrating collective sensemaking and rational decision making in management pedagogy and practice”, Management Decision, Vol 60, No. 11, pp. 3127–3146.

Culmsee, P . and Awati, K . (2012), “Towards a holding environment: building shared understanding and commitment in projects”, International Journal of Managing Projects in Business, Vol. 5 No. 3, pp. 528–548.

Culmsee, P . and Awati, K . (2013), The Heretic’s Guide to Best Practices: The Reality of Managing Complex Problems in Organisations, iUniverse Star, Indianapolis, IN.

Doing Data Science

Huyen, C . (2022), Designing Machine Learning Systems. O’Reilly Media, Inc., Sebastopol, CA.

Kreuzberger, D ., Kühl, N . and Hirschl, S . (2022), Machine Learning Operations (MLOps): Overview, Definition, and Architecture, arXiv preprint arXiv:2205.02302.

Scriven, A ., Kedziora, D ., Musial, K . and Gabrys, B . (2022), “The technological emergence of AutoML; A survey of performant software and applications in the context of industry,” arXiv preprint arXiv:2211.04148.

Žliobaitė, I ., Budka, M . and Stahl, F . (2015), “Towards cost-sensitive adaptation: When is it worth updating your predictive model,” Neurocomputing, Vol. 150, pp. 240–249.

Doing the Right Thing

Chomiak, K . and Miktus, M . (2021), “Harnessing value from data science in business: ensuring explainability and fairness of solutions”. arXiv preprint arXiv:2108.07714.

Culmsee, P . and Awati, K . (2012), “Towards a holding environment: building shared understanding and commitment in projects”, International Journal of Managing Projects in Business, Vol. 5 No. 3, pp. 528–548.

Lee, M and Singh, J . (2021). “The Landscape and Gaps in Open Source Fairness Toolkits. In CHI Conference on Human Factors in Computing Systems (CHI ’21)”, May 8–13, 2021, Yokohama, Japan. ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3411764.3445261

Mäntymäki, M ., Minkkinen, M ., Birkstedt, T . and Viljanen, M . (2022), “Defining organizational AI governance’’, AI and Ethics. https://doi.org/10.1007/s43681-022-00143-x

Mehrabi, N ., Morstatter, F ., Saxena, N ., Lerman, K . and Galstyan, A . (2021), “A survey on bias and fairness in machine learning,” ACM Computing Surveys (CSUR), Vol. 54 No. 6, pp. 1–35.

Mingers, J. C. (1995), “Information and meaning: Foundations for an intersubjective account”. Information Systems Journal, Vol. 5 No. 4, pp. 285–306.

Schneider, J ., Abraham, R ., Meske, C . and Vom Brocke, J . (2022), “Artificial intelligence governance for businesses’’, Information Systems Management, pp. 1–21.

Suresh, H . and Guttag, J . (2021), “Understanding potential sources of harm throughout the machine learning life cycle,” MIT Case Studies in Social and Ethical Responsibilities of Computing (Summer 2021).