

This item was submitted to [Loughborough's Research Repository](#) by the author.
Items in Figshare are protected by copyright, with all rights reserved, unless otherwise indicated.

Science and data: the forbidden fruit

PLEASE CITE THE PUBLISHED VERSION

PUBLISHER

© Patrick Stacey. Published by Achamore Books.

VERSION

AM (Accepted Manuscript)

PUBLISHER STATEMENT

This work is made available according to the conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence. Full details of this licence are available at:
<https://creativecommons.org/licenses/by-nc-nd/4.0/>

LICENCE

CC BY-NC-ND 4.0

REPOSITORY RECORD

Stacey, Patrick K.. 2019. "Science and Data: The Forbidden Fruit". figshare. <https://hdl.handle.net/2134/24467>.

CHAPTER 1:

SCIENCE AND DATA: THE FORBIDDEN FRUIT

By Dr. Patrick Stacey

Loughborough University, UK

p.stacey@lboro.ac.uk

<http://www.drpstacey.com>

Introduction

HelloWorld. Big data analytics and data science have shaken the tree of the corporate world, and lo!, the forbidden fruit has fallen. A fruit that appears to be the juiciest to eat and for which normal cutlery and place-setting will suffice. Sure, there's the promise of its juicy predictive power, real-time monitoring and even data-driven products. But, normal corporate cutlery and place-setting definitely will *not* suffice. Yet, many organizations I consult assume that the same tools (e.g. Hadoop, Google Cloud or IBM Bluemix), environment and process will do the data science trick. This chapter and book lays out why this is not the case; be ever more prepared for change.

The current priority in the industry is to acquire analytics toolsets and data sources and begin work in the 'usual' way. This is a fallacy; a professional big data analyst cum data scientist does not do this; the process requires scientific design. A fundamental understanding of the process and context of big data analytics projects is at least as important as learning the toolset; the priorities of analytics needs revision. A key motivation for this book therefore is to appreciate and understand the process and context of analytics initiatives in terms of organizing and managing them. The book features contributions by advanced researchers, practitioners and data science experts in the form of essays. At the time of writing, there were no such books available that provided this perspective of the analytics pipeline. The management understanding that this book offers is critical to getting the most out of big data analytics initiatives. Some surveys have found that 55% of analytics projects *fail* due to poor process management (Kelley and Kaskade, 2013). There is great potential to be had for businesses and researchers through effective analytics project design and management.

~