Praise for *Halo Data*

'The historical perspective ranging from data as the new "oil" and "gold" is a powerful grounding mechanism. Carruthers and Jackson's analogy to quantum physics and the Halo is sure to carry us all forward to the next seismic shift in data-driven value.'
Iwao Fusillo, Global Head of Data & Analytics – eCommerce, PepsiCo

'Leading the way using storytelling to explain complex data concepts, Carruthers and Jackson have once again written an accessible practitioner's guide to data management. Picking up on the topical challenge of measuring value from our data, Carruthers and Jackson have created a multidimensional metadata framework: a great applied contribution to our field.'
Johanna Hutchinson, Chief Data Officer, BAE Systems

'Packed full of practical case studies as a result of decades of experience, this book manages to stand out, not just because it's an interesting read, but because it achieves something new and unique. A genuinely fresh approach that addresses the inconvenient truth that data and its value are not black and white. It changes, given context, timing and process. This should sit on any data professional's bookshelf and its core idea explored by anyone that genuinely wants to harness the power of data.'
Barry Panayi, Chief Data and Insight Officer, John Lewis Partnership

'As the former Chief Data Officer for Microsoft UK, I have had the privilege of witnessing first-hand the transformative power of data and its ability to shape the future of organisations. *Halo Data* is a comprehensive guide that delves into navigating the intricacies of harnessing the potential of data to drive business success. Carruthers and Jackson explore the history of how we have approached data and turn it upside down in a new and novel way of exploring the components of a holistic data strategy, as well as giving guidance and advice on implementation and execution for your organisation.

This book will serve as your guide, providing you with the knowledge, insights and best practices to navigate the complexities of the data-driven world. As you embark on this journey, I encourage you to

embrace the power of data and its potential to transform your organisation.'
Robin Sutara, Field CTO, Databricks and former Chief Data Officer, Microsoft UK

'A brave new book on treating data as an asset, crafting a data strategy, and creating value from data. A must-read for all data leaders.'
Katia Walsh, Chief Digital Officer, Harvard Business School

'The atomic structure proves a natural proxy to provide a framework to value the data any organisation can gather and use to create competitive advantage, with plenty of guidance on how to realise the opportunities in many scenarios. An approachable and thought-provoking concept from two data industry legends.'
Pete Williams, Director of Data, Penguin Random House UK

Halo Data

Every purchase of a Facet book helps to fund CILIP's advocacy, awareness and accreditation programmes for information professionals.

Halo Data

Understanding and Leveraging the Value of your Data

Caroline Carruthers and Peter Jackson

facet publishing

Published by Facet Publishing
c/o British Library, 96 Euston Road, London NW1 2DB
www.facetpublishing.co.uk

Facet Publishing is wholly owned by CILIP: the Library and Information Association.

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library.

ISBN 978-1-78330-617-6 (paperback)
ISBN 978-1-78330-618-3 (PDF)
ISBN 978-1-78330-619-0 (EPUB)

First published 2023

Typeset from authors' files in 11/14pt Myriad Pro and University Old Style by Flagholme Publishing Services.
Printed and made in Great Britain by CPI Group (UK) Ltd, Croydon, CR0 4YY.

To Dez, Aidan and Jenny

Contents

Figures and Tables

Figures

Tables

About the Authors

Caroline Carruthers

Caroline is an award-winning data expert and co-founder of global data consultancy Carruthers and Jackson. She was one of the first women to take on the role of Chief Data Officer in the UK public sector for Network Rail and has worked across multiple industries to build data capability from the ground up, delivering real change to each organisation around how they value their data as an asset.

Caroline has authored several best-selling books on the role of data in organisations, including the renowned *Chief Data Officer's Playbook* (2017) and the bestselling *Data Driven Business Transformation* (2019). Caroline now consults with public, private and charity sector organisations around the world on how to get the most out of their data.

Caroline was named one of the top 20 inspirational female role models in data by The Female Lead in 2018 and won *Computing*'s Entrepreneur of the Year award in 2020.

Peter Jackson

Peter is co-founder of Carruthers and Jackson and currently Chief Data and Technology Officer at Outra. Previously he was Group Director of Data Sciences at Legal & General and Chief Data Officer of the L&G Investment Management business. Before joining L&G he was Chief Data Officer at Southern Water and prior to that, Head of Data at The Pensions Regulator (TPR), which regulates the pensions and automatic enrolment in the UK. Peter had spent 17 years providing data strategy consultancy across the not-for-profit sector, financial services and FMCG, working with large multinational organisations and blue chip brands. Peter is a specialist in data strategy, data technologies, master

data management strategies, data governance frameworks, GDPR and data science strategies. Peter is the co-author of *The Chief Data Officer's Playbook* (2017) and *Data Driven Business Transformation* (2019). He is an international speaker on data, innovation and business transformation.

Catherine Mandungu

Catherine is an entrepreneur, a revenue operations expert and the founder of a leading revenue operations consultancy Think RevOps. Early in her career she worked for renowned tech giants such as Microsoft, Amazon and Adobe, driving programmatic and operational efficiency by implementing data-driven solutions. She has been recognised as one of the UK's top 100 female entrepreneurs by *The Telegraph* and Natwest. As the founder of Think RevOps, she has helped businesses across multiple technology sectors scale their growth through revenue operations strategies, building data capabilities and digital automation. She is also an advocate for empowering women in business and mentors female founders, as well as women interested in pursuing a career in revenue operations.

Acknowledgements

The list of incredible people who have supported and helped us just continues to grow and grow. We love the data community for its unending support and enthusiasm to share with us and each other, as well as the wonderful characters we are proud to call friends. To everyone who has worked with us at Carruthers and Jackson, past and present, you are all total rock stars especially Ashley Cairns, Andy Lunt and James Miller for your excellent reviewing skills! Our summer school alumni ask us such interesting questions every year and you have no idea how much of the answers you give us fuels our thinking to keep moving forward.

Thanks have to go to our patient and tireless publisher, Facet: Michelle Lau, Pete Baker and our copy-editor, Judith Oppenheimer. You have the patience of saints and we thank you for that because we are both fully aware of our stubbornness and 'interesting' take on deadlines.

From Peter:
Personally I would like to thank my colleagues at Outra who are a constant source of support and great ideas, and keep my ideas grounded in the day-to-day reality of data: Giles Mackey, Alex Young, Caroline Sievers, David Michel and Ryan Ginbey. As ever I want to thank Jenny Heller, whose inspiration gave the title to this book, and who has continued to support my obsession with data and the long hours.

From Caroline:
My thanks have to go to my husband, who is the world's best sounding board and supporter, and to my son: you are the best parts of your dad and me and you make me proud every day, even if you do look blankly

at me when I get excited about data. Thanks also to my family, who always smile knowingly when I say I am never writing a book again.

Introduction

Ever since the mathematician Clive Humby coined the phrase 'Data is the new oil' in 2006, we have all become a bit obsessed with what data is 'like' so as to sell its virtues, to convince more people to be data cheerleaders and work with data as an asset. We have seen the phrase used on numerous occasions: world leaders, business leaders and publications worldwide have picked it up and acted as if it was the most important thing that Humby said. Michael Palmer, writing a blog post in November 2006, stated: 'Data is just like crude. It's valuable, but if unrefined it cannot really be used. It has to be changed into gas, plastic, chemicals, etc. to create a valuable entity that drives profitable activity; so must data be broken down, analyzed for it to have value' (https://ana.blogs.com/maestros/2006/11/data_is_the_new.html).

The point to be understood is that data in its raw form doesn't really do very much. Humby's phrase also portrays that data can be used in many different ways and can be turned into a multitude of different and varied products for us to get some value from it. It sits there full of potential, waiting for us to refine, clean, link, structure and analyse it; basically to unlock it so that we can turn it into a model for predicting when extreme weather will affect us, or how to cope with spikes in demand in our medical services, or how to predict customer or citizen behaviour.

The phrase also highlights that oil and data have some attributes in common. There may be some value in taking our understanding of how we use oil and applying it to data. We can look at how oil as an asset is treated and draw useful parallels for how we can treat data. From understanding what stages oil goes through and how it is treated, we can move on to thinking about how the same principle can be applied to data and the processes data needs to go through in order to be useful;

in other words, the refining process, understanding what it is going to be used for, the preparation phase and so on. The words also bring to mind the engineering and the energy required to convert oil into something useful (think about the complexity and scale of an oil refinery). These are great comparisons to help us move forward our thinking about how we treat data.

Humby's phrase also helps us to be more aware of the negative side of data, and the impact of harmful data leaks and how they could be managed. The pros and the cons of oil can be reflected in the pros and cons of data. Thank you very much, Clive Humby, for that analogy; it did move us all forward. In fact *The Economist* magazine jumped on the bandwagon in 2017 with an article titled 'The World's Most Valuable Resource Is no Longer Oil but Data'. That was more than a decade after he coined the phrase, but they got there eventually.

Analogies can be really helpful in showing us the similarities between things, and this one has really proved its truth because of the number of times it has been picked up and reused. The other side of an analogy is that it may draw our attention to differences between the things we are comparing. In this case the most obvious one is that oil is a finite resource and data is most definitely not. Data can be used, but you still have it, new data is constantly being generated and it is used many, many times by lots of different people and organisations. Data can also be enriched and its value increased. Unlike oil, data isn't a commodity that can be used up.

Now we just want to take a moment here to make sure you are aware that we are in no way, shape or form claiming that Clive Humby doesn't know what he is talking about, he came up with a really useful analogy that made us stop, think and learn. It has generated work for the Economist and Forbes writers and we are sure it will continue to do so. Bearing in mind he came up with this in 2006 when we were still getting our heads around this data thing.

Picking up the mantle of trying to find an analogy that works better has been the source of inspiration for many data folks over the past few years and we have heard some good ones.

Data is the new gold is another analogy that has sprung up. Instead of enterprising individuals going to California to look for gold, they now go to Silicon Valley to create companies to profit from data, acquiring, mining and exploiting it to make their fortunes. It's treated as more valuable than gold and has spawned companies like Google,

Facebook and Twitter who trade in data and their own personal way of mining it to create valuable products and, as a side project have created some of the most valuable companies in the world.

Everyone loves gold: it's shiny, and we all agree it has a value – and to those data cheerleaders among us, data is just as shiny and valuable. This analogy, just like the previous one, relates to a raw asset that needs to be found, mined and, in the case of gold, transformed into a shiny piece of jewellery to adorn your hand. As with gold, the quality of the data is important; you can't make an 18-carat ring from 9-carat gold. In the data world we have the 'garbage in, garbage out' phrase that may well be used till the end of days to promote data governance. Also like gold, data is to be treasured and kept secure.

So is 'data is the new gold' the correct analogy to use? They are both commodities. For both data and gold quality is important. Data needs to be mined and refined like gold. But data in its raw state isn't valuable like gold; data is everywhere, in fact we are being overwhelmed with it in our everyday lives and at work. It is easy to obtain and the supply of it continues to grow. The analogy falls over, and data is not the new gold – it is so much more.

We have heard many versions of the analogy, such as data is the new coal or gas as people try to draw the concept out, always based on the idea of working with a commodity. 'Data is the new plastic' was coined to highlight the great benefits that data can bring when it is correctly harnessed, but also to demonstrate the harmful consequences when it is produced at such a great scale but without an end purpose in mind.

Water gets dragged into the seemingly endless list of metaphors. Unlike oil, which is scarce, water and data are both abundant. A small point but a useful one to make here is that we process oil whereas we purify water – which feels better in terms of what we do to data. Just as water is part of us, making up around 60% of the adult body, personal data is part of us. The analogy is an interesting one because with water quality is important, as it is with data. But it gets more nuanced than that: knowing the quality of the water is important, and water of different qualities is used for different purposes. Just as with water, it is important to know the quality of the data. But even this analogy falls down eventually, for the same reasons as the analogies above.

Another one is 'data is the new soil'. Data supports the growth of business outcomes, and even businesses themselves. Businesses draw

nutrients from the data, just as plants draw nutrients from the soil. This quickly leads to another analogy: 'data is the new sun' – the sun being the source of all energy for businesses to survive, and its rays being abundant and ubiquitous.

Each of these analogies lends itself to furthering our understanding of data and its use.

Data is the new currency. In fact it has even been suggested that data will be the currency of the future. For organisations that trade in data, true data companies, we can see where this comes from. For the rest of us, we're highly sceptical about this one in particular, as we can come up with data until we are blue in the face but I doubt that my local supermarket would exchange my incoherent ramblings for some milk. Going back to the work that Clive Humby was part of, the right data in the form of points might get me that milk but anything else would get us labelled as a bit mad. There is a side note here: crypto-currency is in fact data and does have value and can be used to buy 'real' assets. So crypto isn't really an analogy to help us better understand data.

One of the funniest ones is 'data is the new bacon' because data like bacon can be served in lots of different ways. They both can be a bit polarising in peoples' attitude towards them but with this one we think that we (the data community) are really starting to stretch the analogies a bit thin now.

We have decided to put our heads above the parapet once again, because we think we have been looking at data in the wrong way; not necessarily treating it in the wrong way, but definitely looking at it in the wrong way. The reason why we have moved from one metaphor to another is because they are like the suit that you wore in your twenties, but middle age has ravaged your figure – the suit just doesn't quite fit any more. You can keep going through the wardrobe, trying on suit after suit, or you can accept that things have changed and look for a better way of dressing.

The one thing that everyone can agree on is that data has value, and if the numbers of data companies that have become unicorns are anything to go by, then it has a great deal of value. None of the analogies above addresses the true value of data or provides us with a framework to understand, measure and record value. Going back to Humby's original analogy, oil has a value which is measured in dollars per barrel,

or water in pence per litre. The analogies don't offer us anything for the value of data.

How do we get to the value of data; how do we understand this value; what is the framework for value? Looking at the world in a different way doesn't actually help you if you don't do something with this new knowledge.

A list of numbers – data – without context remains a list of numbers; it is of no real interest there. But if you know that those numbers are the winning lottery numbers and you have the ticket in your hand, then you will be over the moon with excitement and those numbers will actually represent a financial value.

Context is everything when it comes to data!