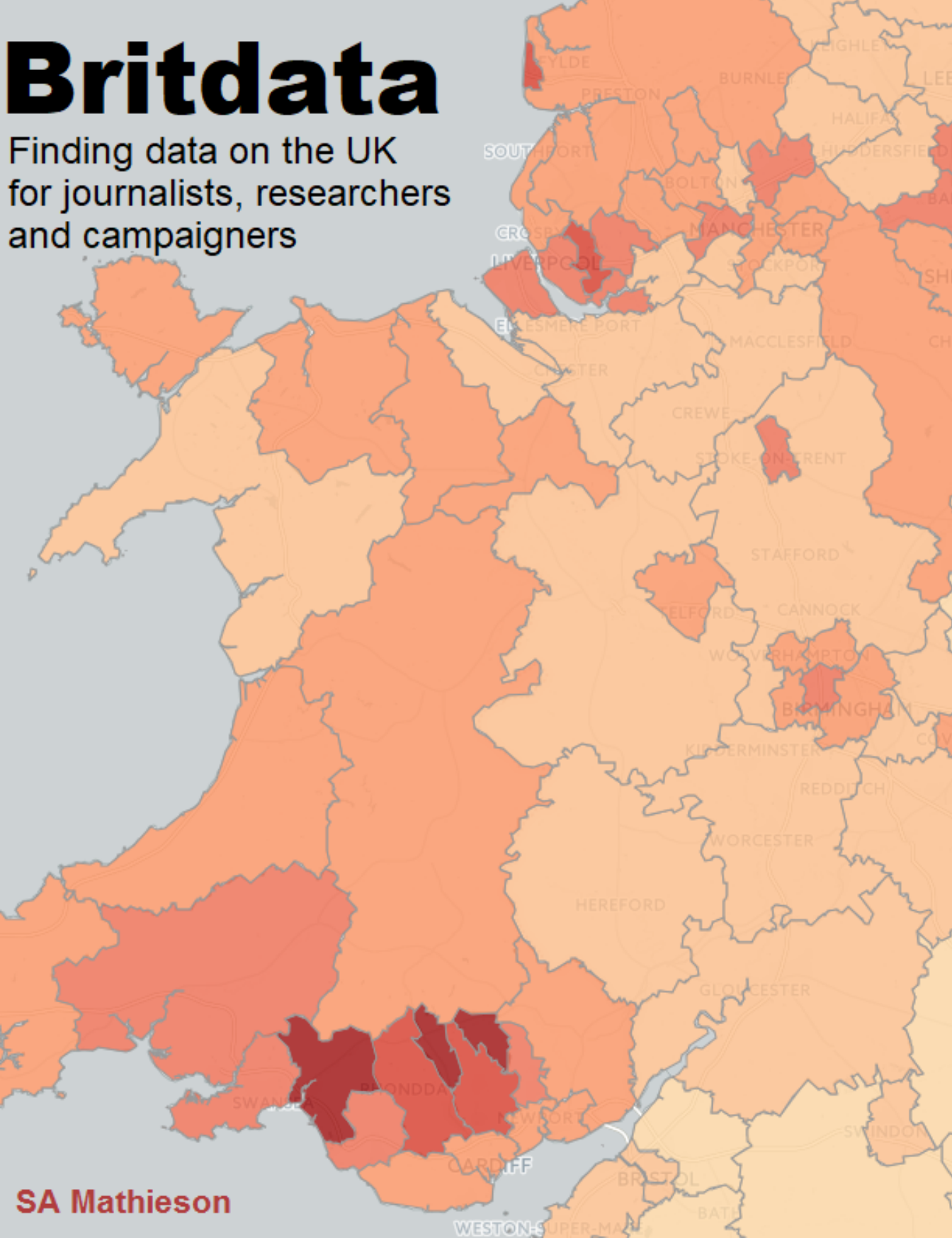


Britdata

Finding data on the UK
for journalists, researchers
and campaigners



SA Mathieson

Britdata

Finding data on the UK for journalists, researchers
and campaigners

SA Mathieson

This book is for sale at

<http://leanpub.com/britdata-uk-data-journalists-researchers-campaigners>

This version was published on 2019-07-19



Leanpub

This is a [Leanpub](#) book. Leanpub empowers authors and publishers with the Lean Publishing process. [Lean Publishing](#) is the act of publishing an in-progress ebook using lightweight tools and many iterations to get reader feedback, pivot until you have the right book and build traction once you do.

© 2018 SA Mathieson

SA Mathieson is a freelance analyst, journalist and editor, specialising in health-care, government and IT. He was launch editor of the Guardian's Healthcare Professionals Network, and writes for publications including The Register and ComputerWeekly.com.

He has run data journalism training for the National Union of Journalists, the Press Association, the Office for National Statistics, City Lit in London and Birmingham City University. He also provides courses on working as a freelancer for the NUJ, as well as one-to-one training in person or online. [More information here](#) or [contact him via email](#). New courses are publicised through [his mailing list](#) and [on Twitter](#).

Contents

Introduction: I spy	1
About this book	2
What is open data?	3
What this book does and doesn't cover	3
Rounding	3
Wikipedia	4
Journalists' Local Authority Directory	4
July 2019 update, thanks and accuracy	5

Introduction: I spy

“Open government is a contradiction in terms. You can be open - or you can have government.” *Sir Humphrey Appleby, permanent secretary of the Department of Administrative Affairs*

Sir Humphrey was the fictional star of the 1980s BBC comedy *Yes, Minister*, but his views reflected the approach of many actually in government. While the UK public sector has opened up somewhat since the 1980s, it still has a way to go.

This partial opening is well illustrated by the UK's signals intelligence service GCHQ.¹ When doing work experience at the old Department of Trade and Industry in the late 1980s, the civil servant I was shadowing told me “that doesn't exist” as we drove past GCHQ's large fenced-off site on the edge of Cheltenham. Details of its work had been revealed by journalists Duncan Campbell and Mark Hosenball in *Time Out* in May 1976, but it took until 1983 for the government even to admit it existed.

Quarter of a century later, GCHQ has [its own website](#), press office and [official puzzle book](#). But unsurprisingly, it remains far from open. It isn't covered by the Freedom of Information Act, introduced by Tony Blair's government and later much regretted by him. It doesn't publish open data - defined below - despite having a number of ways to access everyone else's... within, it can reasonably argue, a strict legal framework.

[Page 84 of the official annual report on the UK intelligence services published by Parliament's intelligence and security committee](#) reveals it had 5,806 staff in March 2015, but p64 of the same document shows that its share of the annual £2.868bn single intelligence account remains a secret - stars indicate where this data has been removed for the public version.

And yet its finances are not a complete mystery. Its new HQ, built on that edge of Cheltenham site, will have cost a total of £1.9815 billion by the time it is paid for in 2030-31; and in the 2015-16 financial year it billed the Home Office

¹GCHQ stand for Government Communications Headquarters, although this doesn't accurately reflect what the organisation does, which is to intercept and decode communications. So it is generally known by the initials.

£170,000 for IT and telecoms services for the Office of Security and Counter-Terrorism and the UK Border Force, [reported here](#).

Some details of its internal security are also public knowledge, such as there being three levels of access to the tool used by staff to access data on telephony and internet data, excitingly named level 1, 1+ and 2, and in 2015 it reported three errors - not bad given MI6 confessed to 11 and MI5 managed 67, [reported here](#).

All of these stories relied on openly-published data of one kind or another. The cost of GCHQ's offices comes from annually-updated data on private finance initiative (PFI) contracts, while the Home Office bill comes from that department's monthly spending data.

The details of how staff access data comes from a legal statement released through an investigatory powers tribunal case against the intelligence and security agencies and the error data comes from the annual report of the intelligence services commissioner.

About this book

Even with a necessarily-secretive organisation like GCHQ, a wealth of data is available to us, hidden only by the fact you have to know where to look for it. The aim of this book is to point out good places to look, particularly for journalists but also for campaigners, researchers and others. While it doesn't aim to be comprehensive, it does aim to highlight a wide range of reliable sources.

The first section aims to provide a guide to the many sources of data on Britain and its public sector which will help you find stories or carry out research with them, broadly divided into open data about the UK; data on national government; data on locally-based government services; data generated by the democratic process; Freedom of Information; other sources; and some tips on dealing with data. The second section provides specific data on local authority areas of the UK.

What such data will not do is provide material on a plate; you can expect to have to dig around extensively, work things out for yourself and see promising leads fizzle out.

But there are advantages, too. The data is usually published by the government itself, meaning any government spokesperson will have trouble denying what

you find. And the fact you'll usually have to do a fair bit of work means that you will often have the results for yourself.

What is open data?

Open data does what it says on the tin. A definition provided by [Open Knowledge International](#) says that "Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness)."

For journalists, writers and researchers it means data you can use as you wish, except you may need to say where it comes from - which you should probably do anyway. Governments including the UK have adopted the release of open data as a way of improving transparency... sometimes undermined by their failure to release some of the more interesting stuff.

What this book does and doesn't cover

This book looks at data beyond that covered by the description 'open data'. I will also cover Freedom of Information and ways to find data in parliamentary records, organisational reports and legal proceedings. As the title suggests, this book covers the United Kingdom, with a few references to international sources of data.

This book will not attempt to teach you how to use spreadsheet software, although there are a few specific tips on specific functions. Examples will use Microsoft Excel, although in most cases the same techniques will work in the free open source package OpenOffice and Google's free online Sheets system.

Rounding

In this book I will unless stated otherwise use data as it is published in spreadsheets, or whatever gets nearest to the raw data. If you use it in further calculations, it makes sense to start with numbers as accurate as possible.

However, this doesn't mean that the Office of National Statistics actually believes that the population of Hertfordshire on 30 June 2018 was exactly 1,184,365, or that this is a sensible number to quote. The ONS calls these figures

estimates and when it quotes them it usually rounds them to the nearest thousand. It's a good idea to do something similar when writing an article or a report, by using two, three or four significant figures.

Such rounding would mean writing that Hertfordshire's estimated population in June 2018 was 1.2m (two significant figures, rounding the last figure up if the following one is 5 or above), 1.18m (three significant figures) or 1.184m (four significant figures).

The organisation you work for may have a style guide which recommends levels of accuracy; I would usually go for three or four significant figures if I have confidence in the data and two or three if less. But this is something to do right at the end of the process, and only to the versions of numbers to be published.

Wikipedia

You might be asking, what is the point of looking for data when it's all on Wikipedia? Because it's not all there, and even when it is it may well be out of date or inaccurate. As an experiment, I once tried adding two zeros to the geographical size of West Oxfordshire District Council on its Wikipedia entry, which would make it roughly the size of Ireland.

I wasn't sure it would be possible to change such an obviously fixed thing as the size of an area of land with no coastline; it was. Then I assumed my change would be spotted and reversed rapidly by a Wikipedia editor; it wasn't, until I discussed it as part of a data journalism course at City Lit six weeks later. One of the students shrunk West Oxfordshire back to its actual size. There are more examples of Wikipedia fictions [here](#).

It's not difficult to get the latest version of data from the original source, and this book aims to make it even easier. Wikipedia is very useful for finding sources of data through the links in its footnotes, but anyone relying on it as a primary source is taking a risk. Remember that it relies heavily on the media, in the rather touching assumption that media professionals obtain and check facts from somewhere more reliable than Wikipedia itself.

Journalists' Local Authority Directory

This book was written at the same time as I was working to set up the Journalists' Local Authority Directory, which contains data covering all the top-

tier local authorities in the UK. A limited extract of the same data is published in mini-profiles at the end of this book, but there is far more available through the online version.

The National Union of Journalists is publicising access to its members in July 2019 - look out for an email with the log-in details. If you are a member of the Chartered Institute of Journalists or the Society of Authors you can get free access to the directory; contact your union for details.

The directory project was funded by the Journalists' Copyright Fund, which is administered by the National Union of Journalists for the three unions.

July 2019 update, thanks and accuracy

This is the third update of Britdata, to include population estimates published in June 2019.

As this is an e-book, I will update it for accuracy and when new data is published. If you bought this through Leanpub or from SAMathieson.com, you will automatically receive an email with the new version.

However, while I have taken care in compiling and processing this data I cannot guarantee its accuracy. It is always best to go back to sources, and I provide links to make this as easy as possible.

I am grateful to Paul Bradshaw, Graeme Bruce, Fran Ethievent, John Kohut and Naomi Goldsmith, who have looked at drafts of this book and made a number of helpful suggestions. However, any mistakes are mine. Please let me know about them via directory@psin.uk.