

Edited by Liliana Bounegru and Jonathan Gray



THE DATA JOURNALISM HANDBOOK

Towards a Critical Data Practice

Amsterdam
University
Press

A
X
U
X
P

Amsterdam
University
Press

“This is a stellar collection that spans applied and scholarly perspectives on practices of data journalism, rich with insights into the work of making data tell stories.”

– Kate Crawford, New York University + Microsoft Research New York; author of *Atlas of AI*

“Researchers sometimes suffer from what I call journalist-envy. Journalists, after all, write well, meet deadlines, and don’t take decades to complete their research. But the journalistic landscape has changed in ways that scholars should heed. A new, dynamic field—data journalism—is flourishing, one that makes the boundaries between our fields less rigid and more interesting. This exciting new volume interrogates this important shift, offering journalists and researchers alike an engaging, critical introduction to this field. Spanning the globe, with an impressive variety of data and purposes, the essays demonstrate the promise and limits of this form of journalism, one that yields new investigative strategies, one that warrants analysis. Perhaps new forms of collaboration will also emerge, and envy can give way to more creative relations.”

– Wendy Espeland, Northwestern University; co-author of *Engines of Anxiety: Academic Rankings, Reputation, and Accountability*

“It is now established that data is entangled with politics and embedded in history and society. This bountiful book highlights the crucial role of data journalists as users and critics of data, and in facilitating public engagement and discussion around it.”

– Emmanuel Didier, Ecole Normale Supérieure; author of *America by the Numbers: Quantification, Democracy, and the Birth of National Statistics*

“This magical multitudinous book is an experiment that will shape the future of critical data journalism.”

– Celia Lury, University of Warwick; author of *Problem Spaces: How and Why Methodology Matters* and co-editor of *Inventive Methods: The Happening of the Social*

“Profound and practical, this sparkling collection engages the topic of data journalism with rich insights into the nature of numbers in the news. It explores both the hows and whys of developing this emergent field through a wonderful range of essays spanning many domains and countries and is of great interest for those producing, consuming and studying news stories based on data.”

– Geoffrey C. Bowker, University of California, Irvine; co-author of *Sorting Things Out: Classification and Its Consequences* and author of *Memory Practices in the Sciences*.



Amsterdam
University
Press

“This book is an impressive feat. Bounegru and Gray have put together a truly global and diverse collection that greatly enriches our understanding of the politics of data and what it means for journalism. In the wake of the Covid-19 pandemic, this contribution is more important than ever.”

– Lina Dencik, co-founder of the Data Justice Lab, Cardiff University; co-author of *Digital Citizenship in a Datafied Society*

“Ostensibly focused on data journalism, this handbook is so much more, providing an overarching analysis of much of the emerging field of critical data studies. Journalists and others interested in how to assemble, work with, make sense of, apply, and critically reflect on data and their uses will revel in the extensive theoretical and practical insights.”

– Rob Kitchin, Maynooth University; author of *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*

“*The Data Journalism Handbook* is an indispensable resource for students, researchers, and journalists who want to understand how data are translated into information, information in knowledge and, ultimately, wisdom. That itinerary all starts with a full comprehension of how data reflect, construct and shape our social reality. Jonathan Gray and Liliana Bounegru have collected and presented an impressive number of illuminating practices in this Handbook.”

– José van Dijck, Utrecht University; author of *The Culture of Connectivity* and *The Platform Society*.

“By providing a wealth of living testimonies from practitioners and academics from different countries, this book gives a rich overview of practices that have become key in contemporary journalism. By cautioning the reader against overconfidence in technology, the main virtue of this book is to give a set of practical insights to help journalists not only to better cooperate with their peers, through computational tools, but also to establish more fruitful relationships with researchers and publics.”

– Sylvain Parasie, Sciences Po; author of *Sociologie d'internet*

“A professional and academic field reaches maturity when it begins to cast a light on itself and becomes capable of self-criticism, asking not only descriptive questions such as *how* we do things, but analytical ones, such as *why* we should—or shouldn't—do those things. The new edition of *The Data Journalism Handbook* signals that we might have reached that point with the use of digital data to do news reporting and inform audiences. The variety, diversity, and depth of the contributions to this collective effort make this book a required reading for beginners and professionals alike.”

– Alberto Cairo, University of Miami; author of *How Charts Lie: Getting Smarter About Visual Information*



Amsterdam
University
Press

The Data Journalism Handbook



Amsterdam
University
Press

Digital Studies

The *Digital Studies* book series aims to provide a space for social and cultural research with and about the digital. In particular, it focuses on ambitious and experimental works which explore and critically engage with the roles of digital data, methods, devices and infrastructures in collective life as well as the issues, challenges and troubles that accompany them.

The series invites proposals for monographs and edited collections which attend to the dynamics, politics, economics and social lives of digital technologies and techniques, informed by and in conversation with fields such as science and technology studies and new media studies.

The series welcomes works which conceptualize, rethink and/or intervene around digitally mediated practices and cultures. It is open to a range of contributions including thoughtful interpretive work, analytical artefacts, creative code, speculative design and/or inventive repurposing of digital objects and methods of the medium.

Series Editors:

- Tobias Blanke, University of Amsterdam
- Liliana Bounegru, King's College London
- Carolin Gerlitz, University of Siegen
- Jonathan Gray, King's College London
- Sabine Niederer, Amsterdam University of Applied Sciences
- Richard Rogers, University of Amsterdam



Amsterdam
University
Press

The Data Journalism Handbook

Towards a Critical Data Practice

*Edited by
Liliana Bounegru and
Jonathan Gray*

Amsterdam University Press



Amsterdam
University
Press

Cover design: Coördesign, Leiden

Lay-out: Crius Group, Hulshout

ISBN 978 94 6298 951 1

e-ISBN 978 90 4854 207 9

DOI 10.5117/9789462989511

NUR 813



Creative Commons License CC BY NC ND (<http://creativecommons.org/licenses/by-nc-nd/3.0>)

©All authors / Amsterdam University Press B.V., Amsterdam 2021

Some rights reserved. Without limiting the rights under copyright reserved above, any part of this book may be reproduced, stored in or introduced into a retrieval system, or transmitted, in any form or by any means (electronic, mechanical, photocopying, recording or otherwise).

Every effort has been made to obtain permission to use all copyrighted illustrations reproduced in this book. Nonetheless, whosoever believes to have rights to this material is advised to contact the publisher.



Amsterdam
University
Press

Table of Contents

Introduction	11
<i>Jonathan Gray and Liliana Bounegru</i>	

Doing Issues With Data

1. From Coffee to Colonialism: Data Investigations Into How the Poor Feed the Rich	27
<i>Raúl Sánchez and Ximena Villagrán</i>	
2. Repurposing Census Data to Measure Segregation in the United States	31
<i>Aaron Williams</i>	
3. Multiplying Memories while Discovering Trees in Bogotá	34
<i>María Isabel Magaña</i>	
4. Behind the Numbers: Home Demolitions in Occupied East Jerusalem	37
<i>Mohammed Haddad</i>	
5. Mapping Crash Incidents to Advocate for Road Safety in the Philippines	41
<i>Aika Rey</i>	
6. Tracking Worker Deaths in Turkey	44
<i>Pınar Dağ</i>	

Assembling Data

7. Building Your Own Data Set: Documenting Knife Crime in the United Kingdom	49
<i>Caelainn Barr</i>	
8. Narrating a Number and Staying With the Trouble of Value	55
<i>Helen Verran</i>	



9. Indigenous Data Sovereignty: Implications for Data Journalism 65
Tahu Kukutai and Maggie Walter
10. Alternative Data Practices in China 74
Yolanda Jinxin Ma
11. Making a Database to Document Land Conflicts Across India 84
Kumar Sambhav Shrivastava and Ankur Paliwal
12. Reassembling Public Data in Cuba: Collaborations When Information Is Missing, Outdated or Scarce 87
Saimi Reyes Carmona, Yudivián Almeida Cruz and Ernesto Guerra
13. Making Data With Readers at *La Nación* 91
Flor Coelho
14. Running Surveys for Investigations 96
Crina-Gabriela Boros

Working With Data

15. Data Journalism: What's Feminism Got to Do With I.T.? 103
Catherine D'Ignazio
16. Infrastructuring Collaborations Around the Panama and Paradise Papers 109
Emilia Díaz-Struck, Cécile Schilis-Gallego and Pierre Romera
17. Text as Data: Finding Stories in Text Collections 116
Barbara Maseda
18. Coding With Data in the Newsroom 124
Basile Simon
19. Accounting for Methods: Spreadsheets, Scripts and Programming Notebooks 128
Sam Leon



20. Working Openly in Data Journalism	138
<i>Natalia Mazotte</i>	
21. Making Algorithms Work for Reporting	143
<i>Jonathan Stray</i>	
22. Journalism With Machines? From Computational Thinking to Distributed Cognition	147
<i>Eddy Borges-Rey</i>	

Experiencing Data

23. Ways of Doing Data Journalism	157
<i>Sarah Cohen</i>	
24. Data Visualizations: Newsroom Trends and Everyday Engagements	162
<i>Helen Kennedy, William Allen, Martin Engebretsen, Rosemary Lucy Hill, Andy Kirk and Wibke Weber</i>	
25. Sketching With Data	174
<i>Mona Chalabi and Jonathan Gray</i>	
26. The Web as Medium for Data Visualization	182
<i>Elliot Bentley</i>	
27. Four Recent Developments in News Graphics	193
<i>Gregor Aisch and Lisa Charlotte Rost</i>	
28. Searchable Databases as a Journalistic Product	197
<i>Zara Rahman and Stefan Wehrmeyer</i>	
29. Narrating Water Conflict With Data and Interactive Comics	206
<i>Nelly Luna Amancio</i>	
30. Data Journalism Should Focus on People and Stories	211
<i>Winy de Jong</i>	



Investigating Data, Platforms and Algorithms

31. The Algorithms Beat: Angles and Methods for Investigation 219
Nicholas Diakopoulos
32. Telling Stories With the Social Web 230
Lam Thuy Vo
33. Digital Forensics: Repurposing Google Analytics IDs 241
Richard Rogers
34. Apps and Their Affordances for Data Investigations 246
Esther Weltevrede
35. Algorithms in the Spotlight: Collaborative Investigations at *Der Spiegel* 257
Christina Elmer

Organizing Data Journalism

36. The #ddj Hashtag on Twitter 267
Eunice Au and Marc Smith
37. Archiving Data Journalism 274
Meredith Broussard
38. From *The Guardian* to Google News Lab: A Decade of Working in Data Journalism 279
Simon Rogers
39. Data Journalism's Ties With Civic Tech 286
Stefan Baack
40. Open-Source Coding Practices in Data Journalism 291
Ryan Pitts and Lindsay Muscato
41. Data Feudalism: How Platforms Shape Cross-border Investigative Networks 295
Ștefan Cîndea



42. Data-Driven Editorial? Considerations for Working With Audience Metrics 299
Caitlin Petre

Learning Data Journalism Together

43. Data Journalism, Digital Universalism and Innovation in the Periphery 307
Anita Say Chan
44. The Datafication of Journalism: Strategies for Data-Driven Storytelling and Industry–Academy Collaboration 314
Damian Radcliffe and Seth C. Lewis
45. Data Journalism by, about and for Marginalized Communities 331
Eva Constantaras
46. Teaching Data Journalism 338
Cheryl Phillips
47. Organizing Data Projects With Women and Minorities in Latin America 344
Eliana A. Vaca Muñoz

Situating Data Journalism

48. Genealogies of Data Journalism 351
C. W. Anderson
49. Data-Driven Gold Standards: What the Field Values as Award-Worthy Data Journalism 360
Wiebke Loosen
50. Beyond Clicks and Shares: How and Why to Measure the Impact of Data Journalism Projects 370
Lindsay Green-Barber



51. Data Journalism: In Whose Interests? <i>Mary Lynn Young and Candis Callison</i>	379
52. Data Journalism With Impact <i>Paul Bradshaw</i>	388
53. What Is Data Journalism For? Cash, Clicks, and Cut and Trys <i>Nikki Usher</i>	397
54. Data Journalism and Digital Liberalism <i>Dominic Boyer</i>	405
Index	411



Introduction

Jonathan Gray and Liliana Bounegru

Abstract

An introduction to the book as a collective experiment in accounting for data journalism practices around the world, providing an overview of its sections and chapters and twelve challenges for critical data practice.

Keywords: data journalism, critical data practice, critical data studies, digital methods, science and technology studies, Internet studies

Data Journalism in Question

What is data journalism? What is it for? What might it do? What opportunities and limitations does it present? Who and what is involved in making it and making sense of it? This book is a collective experiment responding to these and other questions about the practices, cultures, politics and settings of data journalism around the world. It follows on from another edited book, *The Data Journalism Handbook: How Journalists Can Use Data to Improve the News* (Gray et al., 2012). Both books assemble a plurality of voices and perspectives to account for the evolving field of data journalism. The first edition started through a “book sprint” at Mozilla Festival in London in 2011, which brought together journalists, technologists, advocacy groups and others to write about how data journalism is done. As we wrote in the introduction, it aimed to “document the passion and enthusiasm, the vision and energy of a nascent movement,” to provide “stories behind the stories” and to let “different voices and views shine through” (Gray et al., 2012). The 2012 edition is now translated into over a dozen languages—including Arabic, Chinese, Czech, French, Georgian, Greek, Italian, Macedonian, Portuguese, Russian, Spanish and Ukrainian—and is used for teaching at many leading universities and training centres around the world, as well as being a well-cited source for researchers studying the field.

Bounegru, L. and J. Gray (eds.), *The Data Journalism Handbook: Towards a Critical Data Practice*. Amsterdam: Amsterdam University Press, 2021
DOI 10.5117/9789462989511_INTRO



Amsterdam
University
Press

While the 2012 book is still widely used (and this book is intended to complement rather than to replace it), a great deal has happened since 2012. On the one hand, data journalism has become more established. In 2011 data journalism as such was very much a field “in the making,” with only a handful of people using the term. It has subsequently become socialized and institutionalized through dedicated organizations, training courses, job posts, professional teams, awards, anthologies, journal articles, reports, tools, online communities, hashtags, conferences, networks, meetups, mailing lists and more. There is also broader awareness of the term through events which are conspicuously data-related, such as the Panama Papers, which whistleblower Edward Snowden then characterized as the “biggest leak in the history of data journalism” (Snowden, 2016).

On the other hand, data journalism has become more contested. The 2013 Snowden leaks helped to publicly confirm a transnational surveillance apparatus of states and technology companies as a matter of fact rather than speculation. These leaks suggested how citizens were made knowable through big data practices, showing a darker side to familiar data-making devices, apps and platforms (Gray & Bounegru, 2019). In the United States the launch of Nate Silver’s dedicated data journalism outlet *FiveThirtyEight* in 2014 was greeted by a backlash for its overconfidence in particular kinds of quantitative methods and its disdain for “opinion journalism” (Byers, 2014). While Silver was acclaimed as “lord and god of the algorithm” by *The Daily Show*’s Jon Stewart for successfully predicting the outcome of the 2012 elections, the statistical methods that he advocated were further critiqued and challenged after the election of Donald Trump in 2016. These elections along with the Brexit vote in the United Kingdom and the rise of populist right-wing leaders around the world, were said to correspond with a “post-truth” moment (Davies, 2016), characterized by a widespread loss of faith in public institutions, expert knowledge, and the mediation of public and political life by online platforms which left their users vulnerable to targeting, manipulation and misinformation.¹

Whether the so-called “post-truth” moment is taken as evidence of failure or as a call to action, one thing is clear: Data can no longer be taken for granted, and nor can data journalism. Data does not just provide neutral and straightforward representations of the world, but is rather entangled with politics and culture, money and power. Institutions and infrastructures underpinning the production of data—from surveys to statistics, climate science to social media platforms—have been called into question. At the

1 For a critical perspective on this term, see Jasanoff, S., & Simmet, H. R. (2017). No funeral bells: Public reason in a “post-truth” age. *Social Studies of Science*, 47(5), 751–770.

time of writing, as the COVID-19 pandemic continues to roll on around the world, numbers, graphs and rankings have become widely shared, thematized, politicized and depoliticized—as exemplified by daily circulating epidemiological charts referred to by the now ubiquitous public health strategy to “flatten the curve.” At the same time, the fragility and provisionality of such data has been widely reported on, with concerns around the under-reporting, non-reporting and classification of cases, as well as growing awareness of the societal and political implications of different kinds of data from sources—from hospital figures to research estimates to self-reporting to transactional data from tracing apps. The pandemic has broadened awareness of not just using but also critically reporting on numbers and data.

Thus one might ask of the use of data in journalism: Which data, whose data and by which means? Data about which issues and to what end? Which kinds of issues are data-rich and which are data-poor, and why? Who has the capacity to benefit from data and who doesn't? What kinds of publics does data assemble, which kinds of capacities does it support, what kinds of politics does it enact and what kinds of participation does it engender?

Towards a Critical Data Practice

Rather than bracketing such questions and concerns, this book aims to “stay with the trouble,” as the prominent feminist scholar Donna Haraway (2016) puts it.² Instead of treating the relevance and importance of data journalism as an assertion, we treat this as a question which can be addressed in multiple ways. The collection of chapters gathered in the book strive to provide a richer story about what data journalism does, with and for whom. Through our editorial work we sought to encourage reflection on what data journalism projects can do, and the conditions under which they can succeed. This entails the cultivation of a different kind of precision in accounting for data journalism practice: Specifying the situations in which it develops and operates. Such precision requires broadening the scope of the book to include not just the ways in which data is analyzed, created and used in the context of journalism but also the social, cultural, political and economic circumstances in which such practices are embedded.

The subtitle of this new book is “towards a critical data practice,” and reflects both our aspiration as editors to bring critical reflection to bear on

2 Alluding to this work, Verran's chapter in this book explores how data journalists might stay with the trouble of value and numbers.

data journalism practices, as well as reflecting the increasingly critical stances of data journalism practitioners. The notion of “critical data practice” is a nod to Philip E. Agre’s notion of “critical technical practice,” which he describes in terms of having “one foot planted in the craft work of design and the other foot planted in the reflexive work of critique” (Agre, 1997, p. 155). As we have written about elsewhere, our interest in this book is understanding how critical engagements with data might modify data practices, making space for public imagination and interventions around data politics (Gray, 2018; Gray et al., 2018).

Alongside contributions from data journalists and practitioners writing about what they do, the book also includes chapters from researchers whose work may advance critical reflection on data journalism practices, from fields such as anthropology, science and technology studies, (new) media studies, Internet studies, platform studies, the sociology of quantification, journalism studies, Indigenous studies, feminist studies, digital methods and digital sociology. Rather than assuming a more traditional division of labour such that researchers provide critical reflection and practitioners offer more instrumental tips and advice, we have sought to encourage researchers to consider the practical salience of their work, and to provide practitioners with space to reflect on what they do outside of their day-to-day deadlines. None of these different perspectives exhaust the field, and our objective is to encourage readers to attend to different aspects of how data journalism is done. In other words, this book is intended to function as a multidisciplinary conversation starter, and—we hope—a catalyst for collaborations.

We do not assume that “data journalism” refers to a single, unified set of practices. Rather it is a prominent label which refers to a diverse plurality of practices which can be empirically studied, specified and experimented with. As one recent review puts it, we need to interrogate the “how of quantification as much as the mere fact of it,” the effects of which “depend on intentions and implementation” (Berman & Hirschman, 2018). Our purpose is not to stabilize how data journalism is done, but rather to draw attention to its manifold aspects and open up space for doing it differently.

A Collective Experiment

It is worth briefly noting what this book is not. It is not just a textbook or handbook in the conventional sense: The chapters do not add up to an established body of knowledge, but are rather intended to indicate interesting directions for further inquiry and experimentation. The book is not just a practical guidebook of tutorials or “how tos”: There are already countless



readily available materials and courses on different aspects of data practice (e.g., data analysis and data visualization). It is not just a book of “behind the scenes” case studies: There are plenty of articles and blog posts showing how projects were done, including interviews with their creators. It is not just a book of recent academic perspectives: There is an emerging body of literature on data journalism scattered across numerous books and journals.³

Rather, the book has been designed as a *collective experiment* in accounting for data journalism practices of recent years and a *collective invitation* to explore how such practices may be modified. It is collective in that, as with the first edition, we have been able to assemble a comparatively large number of contributors (over 70) for a short book. The editorial process has benefitted from recommendations from contributors during email exchanges. A workshop with a number of contributors at the International Journalism Festival in Perugia in 2018 provided an opportunity for exchanges and reflection. A “beta” version of the book has been released online to provide an opportunity to publicly preview a selection of chapters before the printed version of the book is published and to elicit comments and encounters before the book takes its final shape. Through what could be considered a kind of curated “snowball editorial,” we have sought to follow how data journalism is done by different actors, in different places, around different topics, through different means. Through the process we have compiled and trawled through many shortlists, longlists, outlets and data sets to curate different perspectives on data journalism practices. Although there were many, many more contributors we would have liked to include, we had to operate within the constraints of a printable book, as well as giving voice to a diversity of geographies, themes, concerns and genders.

It is experimental in that the chapters provide different perspectives and provocations on data journalism, which we invite readers to further explore through actively configuring their own blends of tools, data sets, methods, texts, publics and issues. Rather than inheriting the ways of seeing and ways of knowing that have been “baked into” elements such as official data sets or social media data, we encourage readers to enrol them into the service of their own lines of inquiry. This follows the spirit of “critical analytics” and “inventive methods” which aim to modify the questions which are asked and the way problems are framed (Lury & Wakeford, 2012; Rogers, 2018). Data journalism can be viewed not just in terms of how things are *represented*, but in terms of how it organizes *relations*—such that it is not just a matter of producing data stories (through collecting, analyzing, visualizing and narrating data), but also attending to who and what these stories bring together (including

3 https://www.zotero.org/groups/data_journalism_research



audiences, sources, methods, institutions and social media platforms). Thus we may ask, as Noortje Marres recently put it: “What are the methods, materials, techniques and arrangements that we curate in order to create spaces where problems can be addressed differently?”⁴ The chapters in this book show how data journalism can be an inventive, imaginative, collaborative craft, highlighting how data journalists interrogate official data sources, make and compile their own data, try new visual and interactive formats, reflect on the effects of their work, and make their methods accountable and code re-usable. If the future of data journalism is uncertain, then we hope that readers of this book will join us in both critically taking stock of what journalism is and has been, as well as intervening to shape its future. As with all works, the success, failure and ultimate fate of this book-as-experiment ultimately lies with you, its readers, what you do with it, what it prompts and the responses it elicits.

The cover image of this book is a photograph of Sarah Sze’s *Fixed Points Finding a Home* in the modern art museum Mudam Luxembourg, for which we are most grateful to the artist, her gallery and the museum for their permission to reproduce.⁵ While it might not seem an obvious choice to put a work of sculpture on the cover of a book about journalism, we thought this image might encourage a relational perspective on data journalism as a kind of curatorial craft, assembling and working with diverse materials, communities and infrastructures to generate different ways of knowing, narrating and seeing the world at different scales and temporalities. Rather than focusing on the outputs of data journalism (e.g., with screenshots of visualizations or interactives), we wanted to reflect the different kinds of processes and collectives involved in doing journalism with data. Having both serendipitously encountered and been deeply absorbed by Sze’s exhibitions at the Mudam, Venice Biennale, ZKM, the Tate and beyond, we thought her work could provide a different (and hopefully less familiar) vantage point on the practice of data journalism which would resonate with relational perspectives on information infrastructures and “data assemblages.”⁶ Her installations

4 A question that Noortje Marres asked in her plenary contribution to EASST 2018 in Lancaster: <https://twitter.com/jwyg/status/1023200997668204544>

5 Sarah Sze, *Fixed Points Finding a Home*, 2012 (details). Mixed media. Dimensions variable. Mudam Luxembourg Commission and Collection. Donation 2012—Les Amis des Musées d’Art et d’Histoire Luxembourg. © Artwork: Sarah Sze. Courtesy the artist and Victoria Miro. © Photo: Andrés Lejona/Mudam Luxembourg.

6 For relational perspectives on data infrastructures see, for example, the seminal work of Susan Leigh Star: Star, S. L., & Ruhleder, K. (1996). Steps toward an ecology of infrastructure: Design and access for large information spaces. *Information Systems Research*, 7, 111–134; Star, S. L. (1999). The ethnography of infrastructure. *American Behavioral Scientist*, 43, 377–391. For more recent work on “data assemblages,” see, for example: Kitchin, R. (2014). *The data revolution:*

embody a precise and playful sensibility towards repurposing found materials that visually paralleled what we were hoping to emphasize with our editorial of different accounts of data journalism for the book. Bruno Latour recently wrote that Sze’s approach to assembling materials can be considered to affirm “compositional discontinuities” (Latour, 2020) —which sits well with our hopes to encourage “critical data practice” and to tell stories both *with* and *about* the diverse materials and actors involved in data journalism, as we discuss further below, as well as with our editorial approach in supporting the different styles, voices, vernaculars and interests of the chapters in this book.

An Overview of the Book

To stay true to our editorial emphasis on specifying the setting, we note that the orientation of the book and its selection of chapters is coloured by our interests and those of our friends, colleagues and networks at this particular moment—including growing concerns about climate change, environmental destruction, air pollution, tax avoidance, (neo)colonialism, racism, sexism, inequality, extractivism, authoritarianism, algorithmic injustice and platform labour. The chapters explore how data journalism makes such issues intelligible and experienceable, as well as the kinds of responses it can mobilize. The selection of chapters also reflects our own oscillations between academic research, journalism and advocacy, as well as the different styles of writing and data practice associated with each of these.

We remain convinced of the generative potential of encounters between colleagues in these different fields, and several of the chapters attest to successful cross-field collaborations. As well as exploring synergies and commonalities, it is also worth noting at the outset (as astute readers will notice) that there are differences, tensions and frictions between the perspectives presented in the various chapters, including different histories and origin stories; different views on methods, data and emerging technologies; different views on the desirability of conventionalization and experimentation with different approaches; and different perspectives on what data journalism is, what it is for, its conditions and constraints, how it is organized and the possibilities it presents.

Big data, open data, data infrastructures and their consequences. SAGE; Kitchin, R., & Lauriault, T. (2018). Towards critical data studies: Charting and unpacking data assemblages and their work. In J. Thatcher, A. Shears, & J. Eckert (Eds.), *Thinking big data in geography: New regimes, new research* (pp. 3–20). University of Nebraska Press.



After this introduction, the book starts with a “taster menu” on doing issues with data. This includes a variety of different formats for making sense of different themes in different places—including tracing connections between agricultural commodities, crime, corruption and colonialism across several countries (Sánchez and Villagrán), mapping segregation in the United States (Williams), multiplying memories of trees in Bogotá (Magaña), looking at the people and scenes behind the numbers for home demolitions in occupied East Jerusalem (Haddad), mobilizing for road safety in the Philippines (Rey) and tracking worker deaths in Turkey (Dağ). The chapters in this section illustrate a breadth of practices from visualization techniques to building campaigns to repurposing official data with different analytical priorities.

The second section focuses on how journalists assemble data—an important emerging area which we have sought to foreground in the book and associated research (Gray et al., 2018; Gray & Bounegru, 2019). This includes exploring the making of projects on themes such as knife crime (Barr) and land conflicts (Shrivastava and Paliwal) as well as accounts of how to obtain and work with data in countries where it may be less easy to come by, such as in Cuba (Reyes, Almeida and Guerra) and China (Ma). Assembling data may also be a way of engaging with readers (Coelho) and assembling interested actors around an issue, which may in itself constitute an important outcome of a project. Gathering data may involve the modification of other forms of knowledge production, such as polls and surveys, to the context of journalism (Boros). A chapter on Indigenous data sovereignty (Kukutai and Walter) explores social, cultural and political issues around official data and how to bring other marginalized perspectives to bear on the organization of collective life with data. As well as using numbers as material for telling stories, data journalists may also tell stories about how numbers are made (Verran).

The third section is concerned with different ways of working with data. This includes with algorithms (Stray), code (Simon) and machines (Borges-Rey). Contributors examine emerging issues and opportunities arising from working with sources such as text data (Maseda). Others look at practices for making data journalistic work transparent, accountable and reproducible (Leon; Mazotte). Databases may also afford opportunities for collaborative work on large investigative projects (Díaz-Struck, Gallego and Romera). Feminist thought and practice may also inspire different ways of working with data (D’Ignazio).

The fourth section is dedicated to examining different ways in which data can be experienced, starting with a look at the different formats that data journalism can take (Cohen). Several pieces reflect on contemporary



visualization practices (Aisch and Rost), as well as how readers respond to and participate in making sense with visualizations (Kennedy et al.). Other pieces look at how data is mediated and presented to readers through databases (Rahman and Wehrmeyer), web-based interactives (Bentley), TV and radio (de Jong), comics (Luna), and sketching with data (Chalabi and Gray).

The fifth section is dedicated to emerging approaches for investigating data, platforms and algorithms. Recent journalism projects take the digital as not only offering new techniques and opportunities for journalists, but also new objects for investigation. Examples of this are Bellingcat and *BuzzFeed News*' widely shared work on viral content, misinformation and digital culture.⁷ Chapters in this section examine different ways of reporting on algorithms (Diakopoulous), as well as how to conduct longer-term collaborations in this area (Elmer). Other chapters look at how to work with social media data to explore how platforms participate in shaping debate, including storytelling approaches (Vo) as well as affinities between digital methods research and data journalism, including how “born digital” data can be used for investigations into web tracking infrastructures (Rogers) as well as about apps and their associated platforms (Weltevrede).

The sixth section is on organizing data journalism, and attends to different types of work in the field which are considered indispensable but not always prominently recognized. This includes how data journalism has changed over the past decade (Rogers); how platforms and the gig economy shape cross-border investigative networks (Cândea); entanglements between data journalism and movements for open data and civic tech (Baack); open-source coding practices (Pitts and Muscato); audience-measurement practices (Petre); archiving data journalism (Broussard); and the role of the #ddj hashtag in connecting data journalism communities on Twitter (Au and Smith).

The seventh section focuses on learning about data journalism as a collaborative process, including data journalism training programmes and the development of data journalism around the world. This includes chapters on teaching data journalism at universities in the United States (Phillips); empowering marginalized communities to tell their stories (Constantaras; Vaca); caution against “digital universalism” and underestimating innovation in the “periphery” (Chan); and different approaches for collaborations between journalists and researchers (Radcliffe and Lewis).

Data journalism does not happen in a vacuum. The eighth and final section focuses on situating this practice in relation to its various social, political, cultural and economic settings. A chapter on the genealogies of data

7 <https://www.buzzfeednews.com/topic/fake-news>, <https://www.bellingcat.com/>

journalism in the United States serves to encourage reflection on the various historical practices and ideas which shape it (Anderson). Other chapters look at how data journalism projects are valued through awards (Loosen); different approaches to measuring the impact of data journalism projects (Bradshaw; Green-Barber); issues around data journalism and colonialism (Young and Callison); whether data journalism can live up to its earlier aspirations to become a field of inspired experimentation, interactivity and play (Usher); and data journalism and digital liberalism (Boyer).

Twelve Challenges for Critical Data Practice

Drawing on the time that we have spent exploring data journalism practices through the development of this book, we would like to conclude this introduction to the book with twelve challenges for “critical data practice.” These consider data journalism in terms of its capacities to *shape relations* between different actors as well as to *produce representations* about the world. Having been tested in the context of our “engaged research-led teaching” collaborations at King’s College London and the Public Data Lab,⁸ they are intended as a prompt for aspiring data journalists, student group projects and investigations, researcher–journalist collaborations, and other activities which aspire to organize collective inquiry with data without taking for granted the infrastructures, environments and practices through which it is produced.

1. How can data journalism projects tell stories *both with and about data* including the various actors, processes, institutions, infrastructures and forms of knowledge through which data is made?
2. How can data journalism projects tell stories about big issues at scale (e.g., climate change, inequality, multinational taxation, migration) while also *affirming the provisionality* and *acknowledging the models, assumptions and uncertainty* involved in the production of numbers?
3. How can data journalism projects account for the *collective character of digital data, platforms, algorithms and online devices*, including the interplay between digital technologies and digital cultures?
4. How can data journalism projects *cultivate their own ways of making things intelligible, meaningful and relatable through data*, without simply uncritically advancing the ways of knowing “baked into” data from dominant institutions, infrastructures and practices?

8 <https://www.kcl.ac.uk/research/engaged-research-led-teaching> and <http://publicdatalab.org/>



5. How can data journalism projects *acknowledge and experiment with the visual cultures and aesthetics that they draw on*, including through combinations of data visualizations and other visual materials?
6. How can data journalism projects make space for *public participation and intervention* in interrogating established data sources and re-imagining which issues are accounted for through data, and how?
7. How might data journalists cultivate and consciously affirm *their own styles of working with data*, which may draw on, yet remain distinct from, areas such as statistics, data science and social media analytics?
8. How can the field of data journalism *develop memory practices to archive and preserve their work*, as well as situating it in relation to practices and cultures that they draw on?
9. How can data journalism projects collaborate around transnational issues in ways which *avoid the logic of the platform and the colony, and affirm innovations at the periphery*?
10. How can data journalism support marginalized communities to use data to *tell their own stories on their own terms*, rather than telling their stories for them?
11. How can data journalism projects develop their own *alternative and inventive ways of accounting for their value and impact in the world*, beyond social media metrics and impact methodologies established in other fields?
12. How might data journalism *develop a style of objectivity which affirms, rather than minimizes, its own role in intervening in the world* and in shaping relations between different actors in collective life?

Words of Thanks

We are most grateful to Amsterdam University Press (AUP), and in particular to Maryse Elliott, for being so supportive with this experimental project, including the publication of an online beta as well as their support for an open access digital version of the book. AUP is perhaps also an apt choice, given that several of the contributors gathered at an early international conference on data journalism which took place in Amsterdam in 2010. Open access funding is supported by a grant from the Netherlands Organisation for Scientific Research (NWO, 324-98-014), thanks to Richard Rogers at the University of Amsterdam.

The vision for the book was germinated through discussions with friends and colleagues associated with the Public Data Lab. We particularly benefited from conversations about different aspects of this book with Andreas



Birkbak, Erik Borra, Noortje Marres, Richard Rogers, Tommaso Venturini, Esther Weltevrede, Michele Mauri, Gabriele Colombo and Angeles Briones. We were also provided with space to develop the direction of this book through events and visits to Columbia University (in discussion with Bruno Latour); Utrecht University; the University of California, Berkeley; Stanford University; the University of Amsterdam; the University of Miami; Aalborg University Copenhagen; Sciences Po, Paris; the University of Cambridge; London School of Economics; Cardiff University; Lancaster University; and the International Journalism Festival in Perugia. Graduate students taking the MA course in data journalism at King's College London helped us to test the notion of "critical data practice" which lies at the heart of this book.

Our longstanding hope to do another edition was both nurtured and materialized thanks to Rina Tsubaki, who helped to facilitate support from the European Journalism Centre and the Google News Lab. We are grateful to Adam Thomas, Bianca Lemmens, Biba Klomp, Letizia Gambini, Arne Grauls and Simon Rogers for providing us with both editorial independence and enduring support to scale up our efforts. The editorial assistance of Daniela Demarchi and Oana Bounegru has been tremendously valuable in helping us to chart a clear course through sprawling currents of texts, footnotes, references, emails, shared documents, version histories, spreadsheets and other materials.

Most of all, we would like to thank all of the data journalism practitioners and researchers who were involved in the project (whether through writing, correspondence or discussion) for accompanying us, and for supporting this experiment with their contributions of time, energy, materials and ideas without which the project would not have been possible. This book is, and continues to be, a collective undertaking.

Works Cited

- Agre, P. E. (1997). Toward a critical technical practice: Lessons learned in trying to reform AI. In G. Bowker, S. L. Star, B. Turner, & L. Gasser (Eds.), *Social science, technical systems, and cooperative work: Beyond the great divide* (pp. 130–157). Erlbaum.
- Berman, E. P., & Hirschman, D. (2018). The sociology of quantification: Where are we now? *Contemporary Sociology*, 47(3), 257–266. <https://doi.org/10.1177/0094306118767649>
- Byers, D. (2014, March 19). Knives out for Nate Silver. *Politico*. <https://www.politico.com/blogs/media/2014/03/knives-out-for-nate-silver-185394.html>



- Davies, W. (2016, August 24). The age of post-truth politics. *The New York Times*. <https://www.nytimes.com/2016/08/24/opinion/campaign-stops/the-age-of-post-truth-politics.html>
- Gray, J. (2018). Three aspects of data worlds. *Krisis: Journal for Contemporary Philosophy*, 1. <https://krisis.eu/three-aspects-of-data-worlds/>
- Gray, J., & Bounegru, L. (2019). What a difference a dataset makes? Data journalism and/as data activism. In J. Evans, S. Ruane, & H. Southall (Eds.), *Data in society: Challenging statistics in an age of globalisation* (pp. 365–374). The Policy Press. <https://doi.org/10.5281/zenodo.1415450>
- Gray, J., Chambers, L., & Bounegru, L. (Eds.). (2012). *The data journalism handbook: How journalists can use data to improve the news*. O'Reilly Media.
- Gray, J., Gerlitz, C., & Bounegru, L. (2018). Data infrastructure literacy. *Big Data & Society*, 5(2), 1–13. <https://doi.org/10.1177/2053951718786316>
- Haraway, D. J. (2016). *Staying with the trouble: Making kin in the Chthulucene*. Duke University Press. <https://doi.org/10.1215/9780822373780>
- Latour, B. (2020) Sarah Sze as the sculptor of Critical Zones. In B. Latour & P. Weibel (Eds.), *Critical zones: The science and politics of landing on earth* (pp. 158–159). The MIT Press.
- Lury, C., & Wakeford, N. (Eds.). (2012). *Inventive methods: The happening of the social*. Routledge.
- Rogers, R. (2018). Otherwise engaged: Social media from vanity metrics to critical analytics. *International Journal of Communication*, 12, 450–472. <https://dare.uva.nl/search?identifier=e7a7c11b-b199-4d7c-a9cb-fdfidd74d493>
- Snowden, E. (2016, April 3). Biggest leak in the history of data journalism just went live, and it's about corruption. Twitter. <https://twitter.com/Snowden/status/716683740903247873>

About the Authors

Liliana Bounegru (@bb_liliana) and **Jonathan Gray** (@jwyg) are Lecturers at the Department of Digital Humanities, King's College London, co-founders of the Public Data Lab, and Research Associates at the Digital Methods Initiative, University of Amsterdam. They contributed equally as editors of this volume.

