

# Table of Contents

Preface .....	xi
<b>1. Introduction .....</b>	<b>1</b>
What Is Data Journalism?	2
Why Journalists Should Use Data	3
Why Is Data Journalism Important?	6
Filtering the Flow of Data	6
New Approaches to Storytelling	6
Like Photo Journalism with a Laptop	6
Data Journalism Is the Future	6
Number-Crunching Meets Word-Smithing	7
Updating Your Skills Set	7
A Remedy for Information Asymmetry	7
An Answer to Data-Driven PR	7
Providing Independent Interpretations of Official Information	8
Dealing with the Data Deluge	8
Our Lives Are Data	9
A Way to Save Time	9
An Essential Part of the Journalists' Toolkit	9
Adapting to Changes in Our Information Environment	10
A Way to See Things You Might Not Otherwise See	10
A Way To Tell Richer Stories	11
Some Favorite Examples	11
Do No Harm in the Las Vegas Sun	11
Government Employee Salary Database	12
Full-Text Visualization of the Iraqi War Logs, Associated Press	13
Murder Mysteries	15
Message Machine	16
Chartball	16
Data Journalism in Perspective	17
Computer-Assisted Reporting and Precision Journalism	18
Data Journalism and Computer-Assisted Reporting	21



Data Journalism Is About Mass Data Literacy	21
<b>2. In The Newsroom</b>	<b>23</b>
The ABC's Data Journalism Play	24
Our Team	26
Where Did We Get the Data From?	27
What Did We Learn?	27
The Big Picture: Some Ideas	27
Data Journalism at the BBC	28
Make It Personal	29
Simple Tools	29
Mining The Data	30
Understanding An Issue	31
Team Overview	31
How the News Apps Team at the Chicago Tribune Works	32
Behind the Scenes at the Guardian Datablog	34
Data Journalism at the Zeit Online	37
How to Hire a Hacker	41
Harnessing External Expertise Through Hackathons	44
Following the Money: Data Journalism and Cross-Border Collaboration	48
Our Stories Come As Code	51
Kaas & Mulvad: Semi-Finished Content for Stakeholder Groups	54
Processes: Innovative IT Plus Analysis	56
Value Created: Personal and Firm Brands and Revenue	57
Key Insights of This Example	57
Business Models for Data Journalism	58
<b>3. Case Studies</b>	<b>61</b>
The Opportunity Gap	62
A Nine Month Investigation into European Structural Funds	64
1. Identify who keeps the data and how it is kept	65
2. Download and prepare the data	65
3. Create a database	65
4. Double-checking and analysis	66
The Eurozone Meltdown	66
Covering the Public Purse with OpenSpending.org	71
Finnish Parliamentary Elections and Campaign Funding	76
1. Find data and developers	76
2. Brainstorm for ideas	76
3. Implement the idea on paper and on the Web	77
4. Publish the data	78
Electoral Hack in Realtime (Hacks/Hackers Buenos Aires)	78
What Data Did We Use?	79



How Was It Developed?	79
Pros	80
Cons	80
Implications	80
Data in the News: WikiLeaks	81
Mapa76 Hackathon	84
The Guardian Datablog's Coverage of the UK Riots	86
Phase One: The Riots As They Happened	87
Phase Two: Reading the Riots	88
Illinois School Report Cards	88
Hospital Billing	90
Care Home Crisis	92
The Tell-All Telephone	93
Which Car Model? MOT Failure Rates	95
Bus Subsidies in Argentina	96
Who Worked on the Project?	100
What Tools Did We Use?	100
Citizen Data Reporters	101
The Big Board for Election Results	104
Crowdsourcing the Price of Water	106
<b>4. Getting Data</b>	<b>109</b>
A Five Minute Field Guide	110
Streamlining Your Search	110
Browse Data Sites and Services	111
Ask a Forum	113
Ask a Mailing List	113
Join Hacks/Hackers	114
Ask an Expert	114
Learn About Government IT	114
Search Again	115
Write an FOI Request	115
Your Right to Data	116
Wobbling Works. Use It!	121
Case Study 1: Farm Subsidy	121
Case Study 2: Side Effects	122
Case Study 3: Smuggling Death	123
Getting Data from the Web	124
What Is Machine-Readable Data?	125
Scraping Websites: What For?	125
What You Can and Cannot Scrape	126
Tools That Help You Scrape	126
How Does a Web Scraper Work?	127



The Anatomy of a Web Page	127
An Example: Scraping Nuclear Incidents with Python	128
The Web as a Data Source	132
Web Tools	132
Web Pages, Images, and Videos	134
Emails	136
Trends	136
Crowdsourcing Data at the Guardian Datablog	137
How the Datablog Used Crowdsourcing to Cover Olympic Ticketing	139
Using and Sharing Data: the Black Letter, the Fine Print, and Reality	142
<b>5. Understanding Data</b>	<b>147</b>
Become Data Literate in Three Simple Steps	148
1. How was the data collected?	149
2. What's in there to learn?	150
3. How reliable is the information?	151
Tips for Working with Numbers in the News	152
Basic Steps in Working with Data	153
Know the Questions You Want to Answer	153
Cleaning Messy Data	154
Data May Have Undocumented Features	155
The £32 Loaf of Bread	157
Start With the Data, Finish With a Story	158
Data Stories	159
Data Journalists Discuss Their Tools of Choice	161
Using Data Visualization to Find Insights in Data	165
Using Visualization to Discover Insights	165
Which Tools to Use	170
An Example: Making Sense of US Election Contribution Data	170
What To Learn From This	176
Get the Source Code	176
<b>6. Delivering Data</b>	<b>177</b>
Presenting Data to the Public	178
To Visualize or Not to Visualize?	178
Using Motion Graphics	178
Telling the World	178
Publishing the Data	179
Opening Up Your Data	179
Starting an Open Data Platform	180
Making Data Human	180
Open Data, Open Source, Open News	181
Add A Download Link	181



How Was It?	127
Pros	128
Cons	132
Implications	132
Data in the News	134
Mapa78 Hack	136
The Guardian	136
Phase One	137
ic Ticketing	139
and Reality	142
.....	147
The Tell-All	148
Which Car Model	149
Bus Subsidies	150
Who Worked	151
What Took	152
Citizen Data	153
The Big Board	153
Countdown	154
.....	155
Getting Data	157
A Five Minute	158
Streamlining	159
Browse Data	161
Ask a Former	165
Ask a Mailbox	165
Join Hack	170
n Data	170
Learn About	176
Search About	176
.....	177
Wobbling	178
Case Study	178
Case Study	178
Case Study	178
Getting Data	179
What Is It	179
Scraping	180
What You	180
Tools That	181
How Does	181

Know Your Scope	182
How to Build a News App	182
Who Is My Audience and What Are Their Needs?	183
How Much Time Should I Spend on This?	184
How Can I Take Things to the Next Level?	184
Wrapping Up	185
News Apps at ProPublica	185
Visualization as the Workhorse of Data Journalism	186
Tip 1: Use small multiples to quickly orient yourself in a large dataset	187
Tip 2: Look at your data upside down and sideways	189
Tip 3: Don't assume	189
Tip 4: Avoid obsessing over precision	190
Tip 5: Create chronologies of cases and events	190
Tip 6: Meet with your graphics department early and often	190
Tips For Publication	190
Using Visualizations to Tell Stories	191
Seeing the Familiar in a New Way	192
Showing Change Over Time	193
Comparing Values	194
Showing Connections and Flows	195
Designing With Data	196
Showing Hierarchy	197
Browsing Large Databases	198
Envisioning Alternate Outcomes	199
When Not To Use Data Visualization	200
Different Charts Tell Different Tales	201
Data Visualization DIY: Our Top Tools	206
Google Fusion Tables	207
Tableau Public	208
Google Spreadsheet Charts	210
Datamarket	210
Many Eyes	211
Color Brewer	212
And Some More	212
How We Serve Data at Verdens Gang	212
Numbers	212
Networks	213
Maps	213
Text Mining	215
Concluding Notes	215
Public Data Goes Social	216
Engaging People Around Your Data	218