



# Digital Journalism, Drones, and Automation

The Language  
and Abstractions  
behind the News

Cate Dowd

# Digital Journalism, Drones, and Automation



# Digital Journalism, Drones, and Automation

*The Language and Abstractions behind  
the News*

CATE DOWD

OXFORD  
UNIVERSITY PRESS

# OXFORD

UNIVERSITY PRESS

Oxford University Press is a department of the University of Oxford. It furthers the University's objective of excellence in research, scholarship, and education by publishing worldwide. Oxford is a registered trade mark of Oxford University Press in the UK and certain other countries.

Published in the United States of America by Oxford University Press  
198 Madison Avenue, New York, NY 10016, United States of America.

© Oxford University Press 2020

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, without the prior permission in writing of Oxford University Press, or as expressly permitted by law, by license, or under terms agreed with the appropriate reproduction rights organization. Inquiries concerning reproduction outside the scope of the above should be sent to the Rights Department, Oxford University Press, at the address above.

You must not circulate this work in any other form  
and you must impose this same condition on any acquirer.

Library of Congress Cataloging-in-Publication Data

Names: Dowd, Cate, author.

Title: Digital journalism, drones, and automation : the language and  
abstractions behind the news / Cate Dowd.

Description: New York : Oxford University Press, 2020. |

Includes bibliographical references and index.

Identifiers: LCCN 2019036450 (print) | LCCN 2019036451 (ebook) |

ISBN 9780190655877 (paperback) | ISBN 9780190655860 (hardback) |

ISBN 9780190655884 | ISBN 9780190655891 (epub) | ISBN 9780190098445

Subjects: LCSH: Journalism—Technological innovations—United States. |

Online journalism—United States—History—21st century.

Classification: LCC PN4784.T34 D69 2020 (print) |

LCC PN4784.T34 (ebook) | DDC 071/.3—dc23

LC record available at <https://lccn.loc.gov/2019036450>

LC ebook record available at <https://lccn.loc.gov/2019036451>

1 3 5 7 9 8 6 4 2

Paperback printed by Marquis, Canada

Hardback printed by Bridgeport National Bindery, Inc., United States of America

# Contents

|  |    |
|--|----|
| <i>Acknowledgements</i>  | ix |
| <i>Introduction</i>  | xi |
| 1. Formative digital and online technologies in news and journalism                      | 1  |
| Introduction   | 1  |
| 1.1 Alan Turing essentials: configurability and universal systems                        | 3  |
| 1.2 Early mark-up, taxonomies, and online standards                                      | 6  |
| 1.3 Exploitation of newsfeeds, journalists, and atomised data                            | 8  |
| Conclusion   | 10 |
| 2. The impacts of integrated media systems and social media in news                      | 13 |
| Introduction   | 13 |
| 2.1 Digital media systems in broadcasting and new roles for journalists                  | 15 |
| 2.2 News verification and analytics in the era of social media                           | 20 |
| 2.3 Content Management Systems in news: tunnels for social media                         | 24 |
| Conclusion   | 28 |
| 3. The backend of news as a juxtaposition of data and human costs                        | 32 |
| Introduction   | 32 |
| 3.1 Indexing and semantic code in news, towards a vocabulary for journalism              | 34 |
| 3.2 Data journalism, visualisations, mobile tracking, and the human costs of smartphones | 43 |
| 3.3 Risky sources, conflicting uses of metadata, and protection of journalists           | 48 |
| Conclusion   | 52 |
| 4. Smartphones and social media in reporting the asylum seeker crisis in Europe          | 56 |
| Introduction   | 56 |
| 4.1 Smartphones and social media in the European asylum seeker and refugee crisis        | 57 |
| 4.2 Smartphones and journalism in the field: Europe 2015–2016                            | 59 |
| 4.3 Visualisations of refugee and asylum seeker movements                                | 63 |
| Conclusion   | 64 |
| 5. Decoding search and partnerships across news, journalism, and social media            | 67 |
| Introduction   | 67 |

|  |     |
|--|-----|
| 5.1 Semantic search and Linked Data: Berners-Lee, Open Calais, and Murdoch's leverage          | 69  |
| 5.2 Changing roles in journalism in the era of semantic search, data farming, and social media | 76  |
| 5.3 Search, analytics, and machines that learn: big data and semantic hummingbirds             | 84  |
| 5.4 Partnerships in traffic, celebrities, and referrals: nudging automation                    | 91  |
| Conclusion   | 98  |
| 6. Drone journalism and aviation laws, systems, training, and tech trends                      | 107 |
| Introduction   | 107 |
| 6.1 Remotely Piloted Aircraft (RPA) and aerial vision in broadcast media                       | 109 |
| 6.2 Jurisdictions, documents, laws, and software for RPA                                       | 112 |
| 6.3 UK, EU, and US issues: Brexit to ID tags for tracking drones and tech trends               | 114 |
| 6.4 Training for a UAV controller certificate: first-hand experiences and key issues           | 119 |
| Conclusion   | 126 |
| 7. Game design and the Semantic CAT Method for journalism                                      | 130 |
| Introduction   | 130 |
| 7.1 Navigating game design for a semantic epistemology of journalism                           | 132 |
| 7.2 Participatory and Contextual Design techniques in the Semantic CAT Method                  | 138 |
| 7.3 The scrabble of language via Trigger documents   | 141 |
| Conclusion   | 146 |
| 8. Game spaces and the transformation of language for new journalism systems                   | 150 |
| Introduction   | 150 |
| 8.1 Game spaces, player objects, interactions, and modalities                                  | 152 |
| 8.2 The elicitation of vocabulary in journalism, social media, and public relations            | 160 |
| 8.3 Transforming language using the VerbIT technique, nouns, and logic                         | 166 |
| Conclusion   | 176 |
| 9. Emotions, behaviours, and context-aware systems for journalism                              | 179 |
| Introduction   | 179 |
| 9.1 Emotions and behaviours: towards design and representations                                | 181 |
| 9.2 Algorithms to drive logic and emotions in Finite State Machines                            | 187 |
| 9.3 Robo-Journo and Promo-Robo: synthetic AI players   | 192 |

|  |     |
|--|-----|
| 9.4 Ideas and approaches for context-aware systems to protect journalists    | 197 |
| Conclusion   | 202 |
| 10. Language in data processing and ontology engineering for journalism      | 205 |
| Introduction   | 205 |
| 10.1 The exploitation of natural language in data processing                 | 207 |
| 10.2 Ontology engineering for journalism, not folksonomies                   | 211 |
| 10.3 Inaugural ontologies for journalism, social media, and public relations | 219 |
| 10.4 A meta-model with ontologies for pedantic semantic systems              | 229 |
| Conclusion   | 231 |
| Reflections  | 234 |
| <i>Index</i>   | 237 |



# Acknowledgements

Many thanks to Sarah Humphreville, Bronwyn Geyer, and Hallie Stebbins at Oxford University Press for supporting me through the book writing process. The seed for this book began several years ago from research at the intersections of journalism, social media, crisis communication, and computing. The direction of my research is informed by a rich background in information systems and media studies, with some influences from artificial intelligence.

Thank you to former colleagues and researchers at CSU: Kay Nankervis, Professor Terry Bossomaier, Michael Antolovich, Rod McCulloch, and the late Professor John Carroll. Several long-standing journalists and media producers must also be thanked for sharing their depth of knowledge with me about journalism, especially David Brill, Debbie Whitmont, Chris McGillion, and others, and also Sharon Schoenmaker for her precise knowledge of public relations.

I am grateful to Emeritus Professor John Barnden from the University of Birmingham for the opportunity to realise the Turing Arts symposium, which opened up my mind to the complex world of AI. Thank you to the MECO community at USYD in Sydney for so many intellectual exchanges about the digital era (Jonathon Hutchinson, Tim Dwyer, Gerrard Goggin, Fiona Martin, Penny O'Donnell, Kathy Cleland, and others). Thank you also to Paul Arthur for his insights into digital humanities, and to media associates in the United States, Amara Aguilar at USC Annenberg School for Communication and Journalism, and Erkki Hutahmo and Peter Lunenfeld at UCLA.

A special thank you to several people at the ABC who guided me during my visit to Ultimo in Sydney, especially Mark Maley, Mark Corcoran, Stuart Watt, Shane McLeod, and the NSW news team. Thank you to the JERAA and ANZCA communities in Australia, especially for keeping the discussions alive on important issues in journalism and digital media methods. Also, thank you to UNE in Armidale, NSW. Special thanks also to Tom Morton. This book is dedicated to journalists far and wide who have experienced firsthand the impacts of data on journalism. It is also for anyone wanting to understand more about the backend of online technologies, especially semantic search, civilian drones, big data, and the influences of social media on news and journalism.



# Introduction

## **Stirring up the ontological turbulence in big-data clouds with semantics for journalism**

The next generation of systems and practices in journalism will require knowledge beyond online editing techniques, aggregation, social media flow, and assumptions about fake news. The profession may also want to aim for ethical practices in journalism to be embedded in algorithms for new systems. Engagement in an early design phase may also be useful for scoping reforms for online and social media legislation. However, these pursuits require higher levels of understanding about backend data and online systems, as well as development of formal vocabulary for journalism concepts and practices. This new domain knowledge should also be expressed in ontological models, informed by participatory approaches. Some problems to be addressed include editorial control issues and fair distribution of news stories and other challenges of data and online systems. Other problematic issues to address include the lack of transparency in corporate data-sharing arrangements. The semantic language for future systems for journalism will be distinctly different from the vocabulary and classifications used for online news tags. It will also need to distinguish the vocabulary for social media *things* in context of journalism. Most importantly, the design of new systems will need participatory and semantic design methods that can support the need for high-level knowledge of data and semantic search methods. The influence of social media partnerships in news and backend data sharing are other problem areas.

Data via integrated media systems in news organisations flows onto cloud servers where it is processed with a myriad of methods. These hubs are for the new generation of data sharing, where large volumes of data are sorted and processed at accelerated speeds, for a range of purposes. Cloud servers are now literally the highest levels of digital convergence, other than legislation, and the latter is lagging. This is where data is shared for advertising, social media benefits, and other domain purposes. Integrated media systems bring benefits for global networked news media organisations, but they also enable more monetisation of data via cloud servers.

Atomised and cross-referenced data associated with news is enmeshed in search, social media, and advertising techniques in this new era of IT and analytics partnerships. The impacts on news and journalism include shifts in the

loci of journalism, indeed fragmentation in journalism processes and practices. The evidence for this is also seen in substantial redundancies in journalism on a global scale. Some changes in journalism can be understood by a closer look at corporate data worlds and by understanding various data methods, which leads to a need for reviewing data in journalism. This also opens up considerations for the design of new semantic systems that work for journalism.

Some shifts in journalism practices are consistent with the affordances of on-line and digital technology, in particular the lure of big data and the new opportunities from integrated media systems, but these are not the affordances of journalism. Digital systems can also detract from the necessary depths required of journalists, yet they enable journalists to do many new things. For instance, journalists have adapted to mobile and digital technologies when reporting in the field, using smartphones and computer tablets, and they can produce real-time stories and upload them whilst still in the field. Journalists have also conceived alternative uses for social media apps and online tools, and created an abundance of online communities to support each other, but the technological affordances, and other factors, may have narrowed some visions for best practices in journalism. This is compounded by the problems of fake news and the intersections between news and social media, alongside various verification projects and new online tools. However, there is still a need to steer new systems so that they are better aligned to journalism values.

Contemporary journalists must balance the benefits of online systems with the goals of their profession, and many realise that their tasks are enmeshed in search, data, analytics, and social media. These complex intersections primarily benefit online advertising and traffic, yet they have created some loss of control over the functions of news. Nonetheless, journalists must move forward with the dominant technologies and systems of the day, even as they are pulled in different directions because of data and online technologies. The time is ripe to understand data and semantic based systems at the backend of news and social media, which ultimately may also advance the design of new systems for journalists and journalism.

Amongst the digital media systems and apps used in news, streaming media and aerial filming from civilian drones have spawned changes in reporting. Both streaming media and aerial filming were evident in global reporting on the European asylum seeker and refugee crisis circa 2015, and continue to be used widely. Like some social media, streaming media, can be disposable or it can remain on networks and may be shared according to sharing agreements. Drones have also emerged with layers of online connectivity, but ideas for networks of drones are limited, yet advances in networks of this kind could impact on journalism. Context-aware systems with sensor technologies are pervasive in many science areas, including precision agriculture, and these systems can also inform

new possibilities for journalism. Several related ideas are discussed later in this book.

The design of new systems for journalism would serve multiple purposes, from training rookie journalists to protecting journalists in conflict zones. They may be realised as wearable computers that embrace artificial intelligence and emotion engines, and they should be designed to support the goals of journalism. The design approaches for new systems will need to utilise participatory methods to ensure the inclusion of the core values of journalism, and should satisfy automation expectations that are suitable for journalism. The main values of journalism can be captured via the concepts and vocabulary of the domain. From there, new system ideas, such as those presented in this book, are seeds for the future of digital journalism.

Semantic search systems and AI applied to data via cloud servers produce data intersections across online news and social media, but the boundaries between social media and news data are at times indistinguishable. However, journalism could strive to build and preserve its own forms of meaningful data via an ontological base for its own domain. This will also inform the design of systems from a base of journalism, and requires mapping of ethical and professional issues in journalism. For the benefit of journalism, ontological models of journalism can inform semantic knowledge of the domain. These approaches will suit machine learning purposes, but semantic and participatory methods could also protect the journalism values as formal vocabulary and concepts from the domain are embedded in algorithms and systems. This means the articulations from the participatory design process should aim for the most appropriate language and data for the domain. Such approaches may also ensure that journalism is ready for other emerging technologies, such as a quantum<sup>1</sup> computing, which will create a significant shift away from binary thinking.

It could be said that in the beginning there was journalism, but in 1937 the word was with Alan Turing and he spoke with symbols and binary digits to bring forth the printed word. Turing's world of symbols and bits soon led to digital computing and then artificial intelligence in the 1950s. By the end of the century the printed word in the world of news was in decline and before long online news and journalism were surrounded by data farms and click farms, and automation was amongst us, spreading all kinds of stories and fake news.

Digital journalism and online news have uncanny similarities to 17th century philosophy and the Age of Reason. The traces are evident in tags, semantics, search, and analytics that are aligned to logic and reasoning, but machine reasoning and artificial intelligence in the 21st century distinguishes itself with

<sup>1</sup> Quantum computing can utilise four states simultaneously, compared to binary digital bits of zeros and ones that are only on or off. Quantum bits are called Qbits.

reasoning bound to narrow semantic structures. It draws on multiple indexes, applied maths, and many other abstractions. At the backend of this online world, as one layer of knowledge about data and reasoning peels away, another complex layer is revealed. Automated decisions quickly turn into a flurry of deep learning methods and models, and surprisingly, they are not all incomprehensible.

Deep learning in AI is how a system learns by increments and reasons about *things*<sup>2</sup> that we tend to think of as data, but systems that learn and reason about data are not well understood, and even less is understood about design approaches for such systems. A key focus in this book is to unpack the layers of knowledge from indexes to nodes of data and to show firstly how deconstruction of language forms a critical base of understanding for data in semantic systems, such as those at work in online news media, social media, and search. Understanding these layers is vital to envisage and design new ideas for media systems. They will be shaped by the invention of new data, data flows, data transactions, and many other data and data processing features. Some knowledge of data structures and data-centric methods also need to fit into the new design method proposed in this book, to build future systems for digital journalism.

This book opens up new insights into the relationships between language, logic, and computations already at work in online news. It explains how grammar structures are essential for binding data in a certain way, using a triple structure, which adds contexts to data when it is processed on remote servers. It is language structures that transform data, alongside maths principles, such as graph edges, and vectors for proximity computations. The triple structure of a subject, predicate, and object is used as a base structure for data, and it enables bound data to be cross-referenced with other data. Other language structures and fragments of language are also utilised in the segmentation of data, and in data processing.

Semantic search engines draw on well-defined vocabulary in a domain, such as those defined for online news. Social media, such as Facebook, also use semantic models for text-mining. Once the descriptive language of a domain is well-defined, they can be integrated into algorithms with layers of logic and reasoning. The descriptions and definitions in algorithms correspond to data objects and online addresses, or URLs, that can be fetched via indexes, and various classification systems and schemas. However, logic and reasoning in online systems is also at work across algorithms and multiple datasets. They may be for automated reasoning, machine learning, and AI decisions, which are performed at almost unfathomable speeds, mostly via cloud servers.

Journalism is not well defined for semantic data purposes in the same ways as online news and models at the backend of social media. However, journalists use

<sup>2</sup> The word *thing* is associated with the Internet of Things, but it is also used in ontology engineering to describe things in a given world that have potential to become data.

systems in which the frameworks and vocabularies are defined for ‘news’ terms and concepts intersecting with search and social media tags. These semantic systems are not specific to journalism. For more than a decade, journalists have made reactive changes to online systems and social media, but they could take an ontological stance about dominant systems and data. New ideas could include learning systems for journalism and exploration of automation and artificial intelligence (AI) that benefits journalism. First, journalists could benefit from more focus on articulations of what is in their domain. This could be achieved by a closer look at domains that have already defined vocabularies and concepts, like online news and social media. Those models could inform the possibilities for AI and automation for journalism purposes.

This book introduces a formal body of knowledge for journalism and many of its core values. The methods to build up this body of knowledge include techniques for the elicitation and articulation of language working towards an ontology for journalism. These approaches are starting points to leverage semantic contexts so that they are aligned to the values of journalism. They culminate in an emerging ontology for journalism and starting vocabulary for the domain, alongside intersecting ontologies for social media and public relations. How the domain of journalism further defines itself will be partly up to journalists collaborating with ontology engineers. In that process, journalists should ensure that a participatory design method with a semantic edge is used, such as that conceived and proposed in this book.

The formal vocabulary for journalism implies epistemic intersections with social media, search, advertising, and public relations. Sorting ambiguities and similarities in language will be important in the early design phase and essential for protecting the values of the domain. Some discussion in this book is informed by journalists who participated in exploratory research for the start of a semantic and ontological base for journalism. A grounded sensibility emerged amongst journalists, captured in the simplest description of newsrooms as “rabbit warrens” (Participant-journalist, 2008). The participatory design method conceived by the author, namely the Semantic CAT Method, was intended as a specific fit for potential data with a focus on elicitation and articulation of concepts and vocabulary. It also allowed creativity amongst journalists to bring forth the tacit knowledge of the domain. The method was conceived from a blend of game design methods and semantic approaches.

The impacts of data and automation on journalism are partly due to new media partnerships, especially those that are aligned for the purpose of access to big data and the best AI methods. These alignments include search and social media, further characterised by chasing online traffic and monetary gains through automated advertising. Journalists may want to wedge a foot between

the doors of these partnerships across IT, social media, news media, and search, to understand them, but with some degree of caution for future systems.

A glimpse into the backend of IT worlds also reveals how important language is in driving contemporary data. It is not all computations, but language is subjected to arbitrary abstractions, including applications of trigonometry to words to determine connotations. Some computational data methods online make semantic models introduced in this book look quite tame. Even so, the fragmentation of language for data purposes and the computations on language in the era of data are unprecedented. It remains to be seen if any of these computations and abstractions on language will have further impacts on language outside of systems.

Atomised data in news and big-data partnerships have played a significant part in the fragmentation of journalism. This book suggests that journalism should fight back with ontological-based emotion engines, combined with AI and sensor technologies, and other technologies. It presents new ideas for systems that integrate and automate the values of journalism, rather than those associated with advertising. For more than a decade, cloud servers and big data have served corporate sales and corporate platforms, and then social media, search, and advertising remapped and navigated another big course for big data. Media moguls joined in, even on the back of real estate data, as journalists reluctantly fell down. Others turned into new age librarians and data analysts, feeding tags into a global index of search and advertising. Yet, few have understood the ontological roots of systems, but ontological turbulence in the clouds seems inevitable.

## **Overview of chapters**

### **Configurability of integrated systems and the fragmentation of journalism**

In the first two chapters of this book historic contexts of computing blend with a selection of online technologies that have shaped digital journalism and news. The highlights include ideas by Alan Turing from the 1940s on digital technology. In particular, Turing's ideas for universal and configurable systems and machine intelligence that ultimately informed the development of interoperable devices and applications that extend into the online environment. In the 1980s Tim Berners-Lee set in motion the standards for online technologies and applications, and at the start of the 21st century he put forward ideas for a semantic web. In subsequent years entrepreneurs appropriated semantic methods for different ends. Other formative developments included web standards for the

exchange of data and specific languages for online news, such as RSS.XML for syndication and mark-up for online news, like NewsML.

Contemporary integrated media systems used in the production and publishing of news are new hubs for online connectivity to social media, bringing these two worlds closer than ever. These systems have produced a new layer of technological convergence across social media and news, and they have opened up a tunnel for data sharing at the backend of remote servers. The corollary of these relatively new intersections and remote data sharing is more emphasis on analytics in news and more reliance on social media traffic within news, which impacts on journalism practices.

The fragmentation of journalism is quite literal and evident in the attempt to create a verification industry from a fragment of journalism tasks. This is a story of a media mogul's drive for access to big data, especially social media data, and the use of journalism to add credibility in the pursuit of big data. Storyful was Murdoch's "first acquisition after the split of News Corporation into two publicly traded companies" (News Corp, 2014). For about a year or so, Storyful as a verification business was a credible model for Murdoch's business interests, but this was a venture into corporate data sharing via social media. It may have weakened the loci of journalism, as verification tasks in journalism migrated to the first ever social media news agency. For Murdoch, the fragmentation of journalism was yesterday's news, and within a couple of years of his acquisition of Storyful, the focus for that business shifted from verification to brands. By then, Murdoch's foot was well inside the big-data door and the rendezvous with Zuckerberg was in progress.

### Language and online systems at the backend of news, search, and social media

In chapters 3 and 5 a symbiotic relationship between language and grammar structures at the backend of systems and tags for online news stories emerge. The data systems for online news include semantic models and domain vocabulary that are used whenever editors add tags to stories. These tagging systems bind and cross-reference data according to hundreds of well-defined concepts in news, which are structured in hierarchical and logical ways. For the world of news, as distinct from journalism, these classification models are built from an ontological base for data objects and URLs in news, so that they can be searched, but they can also be processed for other ends. These chapters look closer at data flows across news systems to show how language structures and semantic models inform search optimisation. Chapter 3 also explains how cross-referencing is

used to create data visualisations and the ways that data can be combined with layers of maps and graphics, using various software.

Online search engines depend on many abstractions, including multiple indexes and various types of code. When tags are added to a story from a structured classification system, that tag is an abstraction that can be cross-referenced with other data, and the way it is processed on cloud servers is not transparent, but most editors regard the keyword as just sorting for search purposes. However, data tags can be sorted into clusters of information and used for automated decisions, such as selecting news stories, or to target and match advertising based on user types and behaviours. The flow and transactions in data processing are loops of search and reasoning dependent on semantics and AI, which are data-farming methods. With large volumes of data, or big data, data farming looks for the hidden relationships in data that emerge from the edges and intersections of data.

Access to big data is so lucrative it has spawned new partnerships across social media, IT, and news, which are partly characterised by competition for proximity to big data and the best value from AI methods. For advantages with data, some new partnerships in news have even included large real-estate data, as well as social media data. Sharing agreements tend not to be public information, but privacy issues can still emerge from data sharing, especially where data that is “de-identified can be re-identified” (Dwyer, 2016, p. 26). Where data is used to create profile types and granularity about online behaviours, these may be regarded as covert privacy breaches. Journalism coexists in these data worlds through partnerships and backend agreements, and news media is now in the distant clouds, but it remains a force majeure.

In addition to semantic models being used in online news, social media, such as Facebook, also use structured data, as well as unstructured data. Facebook also engages in deep analysis of text in social media posts, which is not what most people expect happens to their posts at the backend of systems. Even if one is not concerned about deep learning on their social media posts, it should be a concern for journalism, and it is a site for future research.

The impacts of search and data processing may be more profound than the rise of data journalism or metadata tracking of images and mobile devices. Decoding the bigger picture of search, including semantic models and tagging systems in news, provides entry points for understanding. Data in journalism can also be leaked from risky sources, which contrasts with standard methods used by journalists, such as actual reporting from conflict zones. The many facets of data across news and journalism in recent years suggests that data and data systems could be used very differently. This revelation opens up the possibilities for the design of new systems to protect journalists in conflict zones. The design of future systems for journalism with data and semantic features should also have a capacity to protect the values of journalism.

Metatags in mobile devices can be used to track reporters in a range of contexts, but few have bothered to track the exploitation of children as slaves extracting raw materials from mines for the manufacture of components in smartphones. Such abuses are also stark contrasts to the madness of identifying thousands of attributes in a single piece of cheese, nominated by online user tagging, for marketing purposes. Yet, these applications of metadata tags could be appropriated for more useful purposes. In 2012 metadata was also used to track the activities of Julian Assange, yet the protection of journalists in actual conflict zones is still limited. It is yet to be seen if new semantic systems can better protect journalists and the values of journalism, in line with the goals of organisations like the CPJ (Committee to Protect Journalists) (CPJ, 2018).

### Field reporting with social media and drone issues for journalism

Approaches to reporting in the field with social media apps, smartphones, and civilian drones are discussed in chapter 4. The focus is on reporting the European asylum seeker crisis circa 2015, where information flowed in many directions and for many reasons. Online mobile systems were also developed to help asylum seekers and refugees. In particular, Google worked the International Rescue Committee (IRC) to “develop sites with location-based information and translations for essential services across central Europe” (Google Creative Lab, 2015). Asylum seekers also used smartphones and apps, such as WhatsApp, for finding locations and getting messages to relatives, and for finding shelter and safety.

Journalists were enmeshed in the asylum seeker crisis that they had little control over and were direct witnesses of the distress and trauma of many displaced people. Any journalists working in the ambit of trauma needs to be prepared for such situations and related stories from the Dart Center for Journalism and Trauma are discussed in this chapter. Former refugees and media producers also used drones for aerial filming for stories and rescuing people at sea. However, as the asylum seeker crisis worsened, drones were used by authorities to track and stop people from crossing borders.

In chapter 6 the topic of drones resumes with a look at the pioneering work done by the BBC and the ABC (Australia). The discussion extends to aviation regulations, training requirements, and qualifications. It delineates professional barriers for media producers from those of the hobbyists and explains why formal training in organisations is necessary, including safety and legal responsibilities. The jurisdictions and documentation for aviation law and the breadth of knowledge to obtain qualifications for flying and operating drones

are outlined. The topics range from aerodynamics, meteorology, and safety to exemption applications and privacy matters. Civilian drones are also discussed in context of digital disruption and policy highlights in the United Kingdom, the European Union, and the United States. Policy discussions in the future may also need to consider issues associated with networks for drones and the significance of deploying drones in swarm formations for visual and other effects, which may signal more than just aerial filming.

The laws and documentation correspond to CASA training for Private Pilot's Licence theory in Australia and derivative training for flying UAVs or RPAs (Remotely Piloted Aircrafts). Steps to become a UAV controller are discussed from first-hand experiences of the author who became a qualified UAV controller in 2016. A significant change in recent years in flying categories now distinguishes between commercial flying of drones for profit and hobbyists, and categories according to the weight of a drone. In 2019 drone registration discussions emerged in Australia via CASA with considerations for lightweight categories.

In recent years RPA flights for media stories are mostly without incident, and media producers tend to use qualified controllers and operators with a RePL (Remote Pilots Licence) for aerial images, rather than take flight risks, or pursue training for a licence. Since late 2016, if filming in a commercial capacity in Australia, an organisation needs to have a ReOC (Remote Operators Certificate) (CASA, 2018) and the drone controller should have a licence.

New ideas for networks of drones would expand the complex layers of on-line data and data hubs. They would also present challenges for controlling aviation space and would raise yet more surveillance issues and the need for new regulations. Even though aviation space is already well defined and safety and collision avoidance remains the highest priority, civilian drones open up discussions about the functions of space and heterotopias. Journalists may want to consider the functions of data that could emerge with networks for civilian drones, and how that might impact further on journalism.

The configuration of drones to fly in swarm formations during major sporting and entertainment events are stark contrasts to drones used to track asylum seekers on the borders of Europe in 2015. Drones as GPS devices fit into the geo-location world of data, and this means any new network ideas will raise new tracking issues, beyond the concerns of collision avoidance. The global trend for drones is that they will be tracked via registration systems for drones, and some of these systems are still emerging. They are driven by combined efforts between government agencies and large IT corporations. Surveillance at these levels could be used for all kinds of ends and will require as much vigilance and knowledge as that still needed for the ethical use of social media data. How journalism might participate in emerging data and network issues for drones remains to be seen.

## Game design and a new semantic method towards an epistemology of journalism

An ontological approach for the design of new journalism systems can be achieved via game design and participatory methods. The ontology approach assists the design of meaningful semantic data for systems. These may be training systems and other systems that protect journalists and journalism values. They will require an epistemology of journalism and other building blocks, which are seeded in chapters 7 and 8. These two chapters outline a new method, the Semantic CAT Method,<sup>3</sup> for ontological approaches to journalism in the early design phase. The method explores tacit concepts and vocabulary in journalism working towards the transformation of language into explicit forms for the design of future systems for journalism. It draws on game design and participatory design methods.

The Semantic CAT Method utilises a range of Trigger documents which have been used in workshops with journalists and public relations professionals to begin defining journalism as a domain. They include drawing and labelling typical scenarios in journalism and some use of UML modelling diagrams for logic and domain processes.

In the design of new systems ambiguity and connotation issues inevitably emerge as language choices need to be made to describe data objects. Yet, shared meanings about language by consensus, for example amongst journalists, can be a way to settle on meanings. This semantic approach to design provides language with contexts bound to a specific concept in a domain. The approach can help semantic meanings to emerge, but a concept must first be explicit.

Intelligent systems require contextual vocabulary and meaningful language statements. A single statement with three data nodes binds data together and defines a simple relationship across that data and is known as a Triple. It takes a collection of Triple statements to build artificial intelligence and data reasoning, amongst other features. Linked data or data statements may be derived from different indexes and dependent on other systems. New systems for journalism may also be based on context-aware models, and chapter 7 introduces some new approaches to the conceptual design of sensor-based systems to protect journalists in conflict zones.

Chapter 8 begins with drawings and creative lists created by journalists with Trigger documents in the Semantic CAT Method. These documents helped to identify, name, and label key concepts in journalism and public relations.

<sup>3</sup> CAT is an acronym for Creative Agents and Triggers. The CAT Method was conceived in 2008 and used in formative workshops associated with an ARC Linkage Project (LP0775418). See <https://ieeexplore.ieee.org/document/5116578/>.

Participating journalists generated vocabulary lists and potential interactions and messages for gameplay. The participatory method complemented an ontology approach, which works to delineate journalism attributes and processes from those in public relations. The participatory process was also one of elicitation and articulation of concepts and vocabulary that informs the ontological base for journalism. This new knowledge in turn, can inform future semantic systems for the domain. The participatory tasks also involved some logic and process models.

Following participatory workshops, a technique called the VerbIT technique, was developed by the author. It takes the main verb in a language statement from journalism and places the verb in first position of the statement, turning it into an imperative, for action. The technique makes it possible to create an entirely new list of character-building verbs for the ontological base of journalism. Nouns in each domain were also considered by the author as potential classes for future systems.

### Emotions, behaviours, and context-aware systems for journalism

New models and systems for journalism are likely to include emotion and behaviour attributes in context of professional transactions and interactions. They could be context-aware systems that draw on data across a range of devices, sensors, and network systems. In chapter 9 new ideas for integrating behaviour attributes into system design could also serve to reinforce ethical behaviours and protect journalists in dangerous zones. The models and systems introduced were created from domain knowledge across journalism, social media, and public relations. They include algorithmic steps for sub-systems that could be used to design synthetic players in future training systems for journalism. They were designed by the author drawing on insights from participant data, communicated in Finite State Machine (FSM) diagrams for models of “Robo-Journo” and “Promo-Robo” (Dowd, 2013). They open up the possibilities for designing behaviour engines for journalism. Some assurances of ethical practices in future systems can be assumed when using participatory and semantic approaches. However, contextualised emotion systems for journalism will need further evaluation and validation.

## The exploits of language processing and ontology engineering

Automation in journalism is not necessarily concerned with robotic reporters, rather integrated media systems connected to social media, cloud servers, big data, and AI systems present enough new challenges for journalism. Language is carved up, sped up, and shuffled in many new ways in data processing, but many will argue that journalism still has the competing edge when it comes to language. Still, for the future of journalism, it may be necessary to understand how natural language processing and learning algorithms in search have already impacted on journalism.

Beyond dashboard analytics in newsrooms, various techniques are used to negotiate and reason about data and online resources. A selection of data methods and computations, applied to language, are discussed in chapter 10. They are used to demonstrate how language is exploited beyond semantic and ontological approaches. However, ontologies are still important for semantic systems, and they are already used to develop semantic search systems for media content and they support social media<sup>4</sup> systems. They can also be utilised to assist the design of new systems for journalism, because they are ideal for distinctions about future data.

Data processing is based on various architectures for systems that work across structured and unstructured data in the name of machine intelligence, or cognitive computing. The backend data methods discussed in chapter 10 include key features of the Hadoop platform first used by Google and Yahoo to manage big data. This chapter also drills down into the abstractions of ontologies that inform semantic systems and data processing. Several techniques to transform journalism language into ontological representations, using the RDF (Resource Description Framework) and concept graphs in OWL are demonstrated. Three inaugural ontologies created by the author, for journalism, social media, and public relations are also introduced, and a meta-model for a future learning system for journalism.

## References

- CASA. (2018, February 20). *Get your RPA operator's certificate*. Retrieved March 25, 2018, from Civil Aviation Safety Authority: <https://www.casa.gov.au/standard-page/get-your-rpa-operators-certificate>.
- CPJ. (2018). *Committee to Protect Journalists*. Retrieved December 2018, from CPJ: <https://cpj.org>.

<sup>4</sup> Facebook's ontological and semantic method are explained in simple terms via this URL <https://www.facebook.com/notes/facebook-engineering/under-the-hood-the-entities-graph/10151490531588920/>

- Dominguez, A. (2017, January 19). *Robot journalist in China makes debut: writes 300-character article in 1 second*. Retrieved May 2017, from itchepost.com: <http://www.itechpost.com/articles/75916/20170119/robot-journalist-china-makes-debut-writes-300-character-article-1.htm>.
- Dowd, C. (2013). *The scrabble of language towards persuasion: changing behaviors in journalism*. In LNCS, S. Berkovsky, & J. Freyne (Eds.), *Persuasive Technology* (Vol. 7822, pp. 39–50).
- Dwyer, T. (2016). *Convergent Media and Privacy*. New York: Palgrave Macmillan.
- Google Creative Lab. (2015, October 13). *Crisis info hub*. Retrieved July 25, 2016, from Crisis Info Hub: [https://docs.google.com/presentation/d/1dta7JJ2v5f04-tdL\\_nRISc13fJw2RE9hXaJt-zF\\_VgY/edit?pref=2&pli=1#slide=id.ge157135a1\\_0\\_0](https://docs.google.com/presentation/d/1dta7JJ2v5f04-tdL_nRISc13fJw2RE9hXaJt-zF_VgY/edit?pref=2&pli=1#slide=id.ge157135a1_0_0).
- News Corp. (2014, April 24). *Storyful partners with Facebook for launch of FB Newswire*. Retrieved January 5, 2016, from News Corp: <http://newscorp.com/2014/04/24/storyful-partners-with-facebook-for-launch-of-fb-newswire/>
- Participant-journalist. (2008). *Dowd CAT Method workshop*. Bathurst, NSW.

# Digital Journalism, Drones, and Automation



# Formative digital and online technologies in news and journalism

## Introduction

It is not breaking news that data tags and machine reasoning enable online systems to track online behaviours and push advertising. However, these features and related methods have impacted on news and journalism. The affordances of digital technology and contemporary online systems can be understood by firstly reflecting on certain traits of Turing machines and early ideas for machine intelligence. This is partly a story about the migration of instructions in computing and technical annotations that began as mark-up in the margins of the print world, which were then modified for the online world. This trajectory prompts new ways of understanding the evolution of code, stemming from standards for the display of online news to code for the aggregation of online stories and newsfeeds. From the history of mark-up emerges the functions of tags and semantic structures that enable contemporary data sharing and predictive analytics via cloud servers. Understanding some features of code opens up an appreciation and opportunities for criticism of the transformative nature of algorithms that can reason about data and digital resources.

Digital media tools and online systems are characterised by dynamic interoperability across devices and systems that are partly due to the affordances of binary, digital technology. They are also shaped by programming and many human factors. The exchange and syndication of online news stories can now be regarded as formative technologies that changed the world of news and journalism. However, this era was surpassed by systems that began to ‘reason’ about online data, including big data, using semantic structures and natural language processing. This new world of artificial intelligence is partly driven by tagging and keywords processed on cloud servers that feed data farming and data mining. The new methods also depend on machine intelligence and coding innovations, which are largely controlled by social media entrepreneurs, search and IT companies, but media moguls also have invested interests. However, many programming contributions at the backend of online systems have emerged from open source and community developer sites. This community of consensus on the suitability and usability of code can also lead to patents locked in by key players.

Formal standards for online news have taken many years to develop and some are open standards, such as NewsML, which defines layout and other on-line features at the backend of news. Some web standards are used for universal applications, such as RSS.XML (Really Simple Syndication.Extensible Mark-up Language), which is base code for newsfeeds, enabling syndication and aggregation. RSS left its mark on news and journalism with profound impacts on the sharing and redistribution of news stories. However, since then the atomisation of data has changed the notion of sharing stories across sites to sharing segments of data associated with stories and online behaviours which are processed at speed via cloud servers. This has been a distraction for journalism, but remote systems have also produced benefits for some reporters.

In recent years the fragmentation of journalism and a corresponding shift in the loci of journalism suggests that it is time to consider new approaches to on-line systems to protect the values of journalism. For those ends journalists may need to engage in the design of new systems, in particular for the integration of ethics and journalism values into new algorithms and artificial intelligence. This means building a deeper and explicit knowledge base of journalism and understanding of how data and digital media systems are designed. This can begin by understanding some of Alan Turing's ideas for digital systems and machine intelligence.

Firstly, journalists are tasked with community dialogue, through reporting, regardless of the tools and methods they use, but data and systems have pulled journalism in many directions. The evolution from bits and things into atomised data and segments of data, and the semantic layers of data, cannot be ignored. At the intersections of search and news, more so than journalism, there are many abstractions of automation at work, and they have already impacted on journalism. Some of the abstractions can be identified in Turing's early ideas for automation.

Machine intelligence and automation does not need to be manifest as a physical robot. Nor should intelligence and automation be dominated by advertising, search, and social media. Intelligent systems for journalism can be conceived, without aiming for an embodied robot dressed up to be a court reporter. However, knowledge of contemporary online systems, including storage, search, semantics, code, configurability, and machine learning in intelligent systems, is useful. The seeds of Turing's abstract ideas for computing, and evolutionary traits, such as configurability, have produced a new generation of digital convergence, and that presents new challenges for journalism.