Stop The Press! Or Press The Stop?

A preliminary study of the convergence of technology and journalism in the Irish Context and its limitations

Deborah Marino
Maud O Connor

BSc (Hons) Information Systems March 2015
Abstract

In recent years a new term has emerged that indicates how journalism is been influenced by technology and automation: computational journalism.

From an initial review of the literature on computational journalism and the overall impact that automation is having on journalism, the researchers set out to look at how Irish journalism is been impacted by this increase in automation.

The primary method of research consisted of semi-structured interviews that saw the involvement of some of the best positioned individual in Ireland to answer the question:

*Is Irish journalism becoming increasingly automated? And if so, is there a limit to that automation?*

In the absence on similar studies in Ireland, this research became a preliminary study of the emergence and impact of computational journalism in Ireland. And for this reason it necessarily evolved in a broad study with a number of findings that address different areas, like the changing role of the journalist, the training and education needs, the need to create a systemic approach to the journalistic workflow and the legal and copyright issues that may emerge.

A number of future research areas have also been identified.

In summary, the study demonstrated that while the evolution of automation in Ireland is slower compared to countries like the UK or the US, there is a thriving enthusiasm for the prospects that the use of software can bring to information.

This enthusiasm seems to be concentrated in a small group of individuals, some of whom have been interviewed, that as part of a spontaneous community of practice, act as advocates of an improved process to manage information and ultimately to provide the public with the most accurate and reliable information.

In this respect the study showed how the majority of the focus in terms of automation and computational journalism is on the data, which translates into an emergence what is known as Data Driven journalism. In Ireland this is what has become synonymous with computational journalism. Albeit the latter is in fact an umbrella term that encompasses other areas of automation, including for example algorithm writing and the use of social media as a journalistic took.
Acknowledgements

We would like to acknowledge all those who supported us throughout this research project. In particular we would like to thank:

- Diana Wilson, our project supervisor, for her support and guidance throughout the project

- All the lecturers that throughout the last four years have contributed to our learning, in particular Simon McGinnes as module owner for his support

- Our research participants, Pamela Duncan, Peter Feeney, Bahareh Heravi, Gavin Sheridan, Paul Watson and Tricia Sheehy-Skeffington for their time and the incredible wealth of knowledge that they shared with us

- The authors of the works reviewed as part of the research which have provided a level of debate and commentary that would provide an ideal foundation for a larger study than was possible here.

Our families, friends and colleagues for their support. Finally a special thank goes to Max, Mike and Ruadhan, who with their support and patience have made this dissertation possible.
Declaration

A dissertation submitted to the University of Dublin in partial fulfilment of the requirements for the degree of B.Sc. (Hons) Information Systems.

We declare that this dissertation has not been submitted before as an exercise for a diploma or degree at this or any other university. The work contained in this dissertation is entirely our own work.

The School of Computer Science and Statistics, Trinity College Dublin may publish this report electronically.

Trinity College Dublin Library may lend or copy the dissertation upon request. The permission only covers single copies for study purposes, subject to the normal conditions of acknowledgement.

Deborah Marino

Maud O Connor

March 30th 2015
Table of Contents

Abstract ................................................................................................................................. ii
Acknowledgements ............................................................................................................. iii
Declaration .............................................................................................................................. iv
Table of Contents .................................................................................................................... v
List of Figures ........................................................................................................................ viii
List of Tables ........................................................................................................................ viii
Abbreviations ....................................................................................................................... ix
Chapter 1 Introduction ........................................................................................................ 1
  1.1 Research Question ........................................................................................................ 1
  1.2 Rationale ....................................................................................................................... 2
  1.3 Project Scope ................................................................................................................ 2
  1.4 Document Structure ..................................................................................................... 3
Chapter 2 Literature Review ............................................................................................... 4
  2.1 What Is Journalism? ..................................................................................................... 4
    2.1.1 What is a journalist? .............................................................................................. 5
    2.1.2 The Journalism Workflow .................................................................................... 6
  2.2 Journalism and technology ......................................................................................... 7
  2.3 Computational Journalism ......................................................................................... 9
    2.3.1 Data Driven Journalism ....................................................................................... 10
    2.3.2 Automated writing .............................................................................................. 11
    2.3.3 Social Media ....................................................................................................... 12
  2.4 Impacts of Increased technical automation in Journalism ............................................. 13
    2.4.1 Automating journalism ......................................................................................... 13
    2.4.2 Using computation to strengthen Journalistic Principles ..................................... 14
    2.4.3 The limits of Automation in Journalism ............................................................... 15
    2.4.4 Journalist versus Data Scientist – the evolving role of the Journalist ............... 16
  2.5 Computational Journalism in Ireland – the current view ............................................. 17
2.6 Conclusion ......................................................................................................................17

Chapter 3  Methodology ..................................................................................................19

3.1 Introduction ................................................................................................................19

3.1.1 Knowledge Claim ..................................................................................................19

3.1.2 The inductive approach .........................................................................................20

3.1.3 Literature review ....................................................................................................20

3.2 The methods of collection of primary data ...............................................................21

3.2.1 Interviews ..............................................................................................................21

3.2.2 Online course .......................................................................................................23

3.3 Qualitative analysis ...................................................................................................24

3.3.1 Data Analysis .......................................................................................................24

3.4 Secondary data ..........................................................................................................24

3.5 Tools ..........................................................................................................................24

3.6 Limitations and constraints of this research ..............................................................25

3.7 Conclusion ................................................................................................................25

Chapter 4  Findings and Analysis ....................................................................................26

4.1 Introduction ...............................................................................................................26

4.2 How was data analysed? ...........................................................................................26

4.3 Participants ...............................................................................................................26

4.4 The changes in Journalism in Ireland ......................................................................29

4.4.1 Future of journalism .............................................................................................30

4.5 Automation ...............................................................................................................30

4.5.1 Limits of automation ............................................................................................32

4.6 Computational Journalism .......................................................................................33

4.6.1 Data Driven Journalism .........................................................................................33

4.6.2 Automated writing ................................................................................................35

4.7 Education and Roles ................................................................................................36

4.7.1 Education ..............................................................................................................36
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.7.2</td>
<td>Roles</td>
<td>38</td>
</tr>
<tr>
<td>4.8</td>
<td>Regulatory Framework</td>
<td>40</td>
</tr>
<tr>
<td>4.9</td>
<td>Potential Legal Issues</td>
<td>41</td>
</tr>
<tr>
<td>4.9.1</td>
<td>Copyright</td>
<td>41</td>
</tr>
<tr>
<td>4.9.2</td>
<td>Data Protection</td>
<td>42</td>
</tr>
<tr>
<td>4.10</td>
<td>Conclusions</td>
<td>42</td>
</tr>
<tr>
<td>Chapter 5</td>
<td>Conclusion</td>
<td>43</td>
</tr>
<tr>
<td>5.1</td>
<td>Areas of future research</td>
<td>44</td>
</tr>
<tr>
<td>5.1.1</td>
<td>A more comprehensive study of the journalism industry</td>
<td>44</td>
</tr>
<tr>
<td>5.1.2</td>
<td>Digital Archiving</td>
<td>44</td>
</tr>
<tr>
<td>5.1.3</td>
<td>Regulatory and legal gaps</td>
<td>44</td>
</tr>
<tr>
<td>5.1.4</td>
<td>Copyright law</td>
<td>44</td>
</tr>
<tr>
<td>5.1.5</td>
<td>Employment</td>
<td>45</td>
</tr>
<tr>
<td>5.1.6</td>
<td>Automated writing</td>
<td>45</td>
</tr>
<tr>
<td>Appendices</td>
<td></td>
<td>46</td>
</tr>
<tr>
<td>Appendix I</td>
<td>– CSO Census 2011 Table 2</td>
<td>46</td>
</tr>
<tr>
<td>Appendix II</td>
<td>– CSO Census 2011 figures</td>
<td>47</td>
</tr>
<tr>
<td>Appendix III</td>
<td>– Interview Questions</td>
<td>48</td>
</tr>
<tr>
<td>Appendix VI</td>
<td>– List of Data Driven Journalism Courses offered worldwide</td>
<td>49</td>
</tr>
<tr>
<td>References</td>
<td></td>
<td>51</td>
</tr>
</tbody>
</table>
List of Figures

FIGURE 1 TRADITIONAL JOURNALISM WORKFLOW ............................................................... 6
FIGURE 2 DIGITAL JOURNALISM CORE WORKFLOW (ANAND, 2014) .................................... 6
FIGURE 3 INVERTED PYRAMID OF DDJ (BRADSHAW, 2013) ................................................... 7
FIGURE 4 FLOWCHART OF TOOLS AND TECHNIQUES FOR DATA GATHERING (BRADSHAW, 2011) ................................................................. 11
FIGURE 5 REASONS FOR USE OF SOCIAL MEDIA BY IRISH JOURNALISTS (HERAVI ET AL, 2014) ......................................................... 12
FIGURE 6 DETERRENTS FROM USING SOCIAL MEDIA (HERAVI ET AL, 2014) ......................... 13
FIGURE 7 PLOTTING OF PARTICIPANTS ON THE TECHNOLOGY/JOURNALISM SPECTRUM .......................................................... 28
FIGURE 8 MAP OF THE COMPUTATIONAL JOURNALISM COURSES OFFERED WORLDWIDE (PALUMBO, 2013) ........................................ 37

List of Tables

TABLE 1 KEY PARTICIPANTS ................................................................................................... 27
TABLE 2 DATA DRIVEN COURSES ....................................................................................... 50
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI</td>
<td>Artificial Intelligence</td>
</tr>
<tr>
<td>BAI</td>
<td>Broadcast Authority of Ireland</td>
</tr>
<tr>
<td>CAR</td>
<td>Computer Aided Reporting</td>
</tr>
<tr>
<td>CSO</td>
<td>Central Statistics Office Ireland</td>
</tr>
<tr>
<td>CTO</td>
<td>Chief Technology Officer</td>
</tr>
<tr>
<td>DCU</td>
<td>Dublin City University</td>
</tr>
<tr>
<td>DDJ</td>
<td>Data Driven Journalism</td>
</tr>
<tr>
<td>DIY</td>
<td>Do It Yourself</td>
</tr>
<tr>
<td>FOI</td>
<td>Freedom of Information (Acts)</td>
</tr>
<tr>
<td>ID</td>
<td>Identification Code</td>
</tr>
<tr>
<td>IS</td>
<td>Information Systems</td>
</tr>
<tr>
<td>MS</td>
<td>Microsoft</td>
</tr>
<tr>
<td>OS</td>
<td>Operating Systems</td>
</tr>
<tr>
<td>RSS</td>
<td>Rich Site Summary</td>
</tr>
</tbody>
</table>
Chapter 1 Introduction

If it is true that, as Joseph Pulitzer said, “the power to mold the future of the Republic will be in the hands of the journalists of future generations”, then it is in a society’s best interest to ensure that a journalist is empowered to fulfil his or her mission by being able to access the most effective tools, to access data and facts and transform them into information.

In recent years a new concept has arisen – computational journalism – to describe the increasing use of technology in journalism. Broadly it is an attempt to define the way in which technology (networking, algorithms, software, and information management) has become embedded in journalism.

The impacts are seen throughout the journalism workflow, from gathering the news, verification of the sources, contextualisation and even in some instances producing the story output, though as will be seen from the research this is confined to specific news topic for now.

The impacts are seen most commonly in the way in which news is identified and gathered. Software tools scan through online sites looking for trends and links, recommender products help journalists link entities and people to build a community of online sources, and imaging verification software, translation and mapping products aid in the verification layer.

Computational journalism is a many-faceted term that covers components such as the use of algorithms to automate writing, management of social media platforms for news-gathering and data-driven journalism where the principles of data mining and processing are employed to support stories, or provide the sources for stories.

The emergence of technology as a disruptive force has caused ripples of disquiet in an industry that is already evaluating its business model due to the disruption caused by the collaborative nature of web 2.0. There are questions being asked as to whether technology practices can uphold the values of objectivity, transparency and social commentary that are specific to journalism and whether two such different practices can join to empower journalists.

1.1 Research Question

The research question being asked is:

Stop The Press! Or Press The Stop?

A preliminary study of the convergence of technology and journalism in the Irish Context and its limitations

The researchers ask: Is Irish journalism becoming increasingly automated? And if so, is there a limit to that automation?

Will automation, to use the analogy of the printing press, cause widespread disruption in journalism and change the way the news is being produced by journalists?

Or will there be a natural limit to automation in that those processes than can be automated will be, and journalists will continue to play a central and complementary role?
Starting from the review of the literature the researchers set to investigate how Irish journalism has been impacted by automation. The focus of the research is to understand the knowledge gleaned from the literature and see if they are applicable to the Irish journalism industry, either currently or in the future.

1.2 Rationale

What brought the researchers together was a shared interest in the impacts of technology on society. As information systems practitioners they felt that there is a responsibility to understand how technology aids or hinders human endeavours. The topic of automation, and the impacts of increased mechanisation of work practices and processes, is one that is the source of debate currently on the benefits and limitations of removing the human from the labour force.

The researchers felt in order to put boundaries around the scope of the research that a focus on a specific industry would yield clearer results and journalism was chosen because of its important role as an observer and watchdog for society.

Technology and software automation is embedded in practically every aspect of today’s life. It may be argued that for some industries the impact is more visible and predictable (e.g. manufacturing), but for certain industries the way they respond to the impact of technology, and in particular software automation, may not be as obvious. For the researchers this was certainly the case with journalism.

From the initial review of the literature it was clear that Ireland was not yet embracing computational journalism and that any local research on the topic may not yield sufficient data to analyse. However, as the initial proposal phase progressed, green shoots began to appear in the industry, which encouraged the researchers to take a renewed look at exactly what was emerging in the Irish scene.

It was agreed, after consultation, that this would be an opportunity to conduct a preliminary study of computational journalism in Ireland in the hope that any findings and specified areas of future research can contribute to the development of a taxonomy of computational journalism in Ireland as a topic of study.

Initially the study focused on one particular aspect of automation, algorithm writing, which is covered as part of the research. However when the study began to take shape it was clear that a broader study would be more useful to the industry.

“History reveals that the more democratic a society, the more news and information it tends to have.” (American Press Institute, 2015)

Coming from information systems the researchers had an interest in looking at the technology that empowers the journalist and that indirectly contributes to the fulfilment of a more democratic society.

1.3 Project Scope

This research explored the adoption of automation in Irish journalism through the spread of computational journalism. While the research has been quite broad given that it consisted of a preliminary study, some exclusion must be specified in order to contain the scope.

This has necessitated the disregarding of issues of on-going debate on the possible demise of the printed newspaper, the movement of daily news outlets to online sites and the economics of
charging for content on online sites, though the authors acknowledge that they are also integral components of many of the subjects examined here.

The way social media has impacted on journalism, for example creating citizen journalists as bloggers, is not in scope of this research but only a peripheral input to the discussion.

The researchers note that they have not worked in journalism and though they attempted to bridge that gap by completing an online course in data driven journalism as part of the preparation for the study, they are limited by a knowledge gap.

1.4 Document Structure

Chapter 1 – Introduction
The topic of Computational Journalism is introduced and discussed. The research question and its rationale are defined to introduce the reader to the journey that unfolds in the following chapters.

Chapter 2 – Literature Review
The chapter consists of a critical review of the literature that supported the definition of the research question and guided the researchers in the development of the research methodology. The literature review also provides a description of the core components of computational journalism.

Chapter 3 – Methodology
The methodology is described, outlining the philosophical rationale, the methods used to gather data and how the data is analysed.

Chapter 4 – Findings & Analysis
This chapter outlines the analysis applied to the gathered data and presents the findings of the research.

Chapter 5 - Conclusion
This chapter contains a final evaluation of the study and its relevance as a basis for further studies; it proposes areas of future research which developed as sub-themes through the analysis phase.
Chapter 2 Literature Review

2.1 What Is Journalism?

Before reviewing literature on computational journalism the authors felt it was important to review definitions of journalism, particularly in light of the growth of blogging and DIY journalists.

The American Press Institute provides the definition of journalism that most appropriately captures a legacy view of the practice, describing it as “the activity of gathering, assessing, creating, and presenting news and information. It is also the product of these activities”. (American Press Institute, 2015). However, blogging is democratising all concepts of journalism with the rise of the citizen blogger and the vast online community. With the advancements in technology and the ubiquity of mobile phones, anyone can communicate with a global audience resulting in millions of people around the world who are actively observing, recording and communicating events as they unfold (Pavlik, 2000).

As such, legacy definitions of journalism are becoming harder to align to the new digital reality and Knight suggested that any definition of journalism concepts are transitional as journalists and analysts attempt to categorise the conjunction of legacy and digital media (Knight et al., 2008).

Bloggers are not obliged to verify sources and maintain impartiality in the way journalists are, and neither are they obliged to operate under a code of practice. Knight catalogues the disquiet felt by journalists that the dispersal of the ‘news’ across new media might leave many of the core values behind, resulting in a less democratic and principled structure (Knight et al., 2008). Indeed Pavlik posits that in the new reality, where citizens can source and report the news themselves, the need for journalists in their current function is negated. He suggests a new role whereby journalists act as interpreters of the news and provide additional meaning and perspective on the events being reported, rather than focus on the reporting itself (Pavlik, 2000). The authors foresee issues with the diminution of the journalist in this manner as it assumes a degree of structure and uniformity in the manner of the reporting and as the citizen blogger is not constrained by any formal framework, this does not appear likely.

The core tenets of journalism – transparency, accountability, objectivity and liberty - are the values that elevate it beyond the sum of its parts. The American Press Institute has written: “History reveals that the more democratic a society, the more news and information it tends to have” (American Press Institute, 2015) and this connects with the role of journalism as society’s democratic watchdog. Franklin quotes from the Levenson inquiry where Lord Justice Levenson highlighted the importance of journalism in the proper functioning of a state, and conversely that the failure of journalism to fulfil this role impacts all citizens (Franklin, 2014). Arthur Miller, in an interview with the London Observer in 1961, mused, “A good newspaper, I suppose, is a nation talking to itself” and journalism acts as the curator and arbitrator of that conversation.

The authors used this review of journalism, and the pressures on it to tailor its legacy principles to the digital world, as way-markers for a review of the impact of technology. Coming from the information Systems practice, the authors were cognisant of evaluating whether technology empowers the journalist and therefore indirectly contributes to the fulfilment of a more democratic society or whether it is an inhibiting factor.

It was first necessary to look at what is meant by a journalist.
2.1.1 What is a journalist?

The literature did not reveal any simple definition of the journalist. The Oxford English dictionary defines a journalist as “A person who writes for newspapers or magazines or prepares news to be broadcast on radio or television” but that definition seemed dated and limited in the digital age. With the proliferation of mobile technology and the Internet a journalist is no longer tied to the newsroom or the television station.

A new definition must be found, particularly as the Internet throws open the doors of the newsroom and everyone can now be a reporter. Gerlis suggests the “who is a journalist” question must be changed as the answer is everyone (Knight et al., 2008).

Knight warns that simply because the Internet is a global publisher that does not mean that people have the requisite competencies to be journalists without first gaining an understanding of the professional and ethical codes of the industry. He predicts a place for trained journalists amid the blogging noise who can verify and contextualise information and provide leadership for an informed audience (Knight et al., 2008). Carlson goes so far as to say that vocational journalism is even more fundamental than ever due to its tradition of fact checking and objectivity when juxtaposed with DIY Journalism – (Carlson, 2014)

Gerlis suggests that professional journalists must clearly distinguish themselves from amateurs by emphasising the professional principles and codes under which they must practice. The ability to supply content that is verified and trustworthy is a differentiating factor and the onus is on the professional journalist now to maintain an even stricter adherence to standards and training in order to segregate their work from that of the layperson. (Knight et al., 2008).

It appears to the authors that there is a place for both the amateur and the profession in line with the principle that the more information that is available the healthier the democracy. Gerlis states that it is incumbent on journalists to come to an accommodation with citizen journalists to work in harmony (Knight et al., 2008).

In terms of how technology integrates with the work of the journalist, Primo suggested looking beyond defining the journalist solely in terms of their outputs. He proposed that journalism is not solely produced by the human actors in the network but also by the technology networks and tools that are integrated in the process (Primo and Zago, 2014). Like Primo the authors focused on the question “who and what do journalism” (Primo and Zago, 2014).
2.1.2 The Journalism Workflow

The traditional workflow process in journalism can be seen as three complementary steps:

![Figure 1 Traditional Journalism Workflow](image1)

The facts were gathered and ascertained, the story was written and then published in the newspaper and sent out onto the streets and into the shops.

This has shifted with the accelerating movement of journalism to the digital sphere (Jurkowitz, 2015) and an amended workflow could now look more like the following:

![Figure 2 Digital Journalism Core Workflow (Anand, 2014)](image2)

In Figure 1 above, the content gathering first stage might use tools to extract data from social media to get content. The movement to digital journalism has shifted emphasis down the workflow to how online content can generate its own community content and the focus is often on the possibilities to monetise this stage.

The workflow process for data driven journalism, one of the core components of the umbrella term computational journalism, has been formalised to a degree by Paul Bradshaw in the diagram illustrated in Figure 3 and is used in data journalism courses worldwide. The inverted pyramid is based on the inverted pyramid of writing taught in traditional journalism courses.

The initial Compile stage is driven by stories that require data or data that can yield stories. It is the most important stage as it is the point on which all other activities are based, and it may be iterative, as the other stages require a review of the gathered data or indeed new sources. The Clean stage involves transforming the data into a useable format and can be manual or aided by tools such as OpenRefine. The Context stage is associated with checking the data to ensure it implies no bias or lack of objectivity in its gathering. It also involves looking at the data in new ways to see if new questions or stories can be found. The final Combine stage may include linking the data to another source to provide a more complete picture or layering results over maps or other visual means to help the reader understand the story. Finally there is a story to be communicated to the reader, whether visually or narratively or indeed a combination of methods (Bradshaw, 2013).
2.2 Journalism and technology

“Whether it’s the pen or the printing press, journalists have always been beholden to the technology that has carried their message throughout history” (Diakopoulos, 2012).

The authors used the literature to understand the evolution of the relationship of journalism and technology and how technology has become an increasingly integrated part of the methodology, not an accessory (Primo and Zago, 2014).

Primo, echoing Pavlik (Pavlik, 2000), argued for a parallel evolutionary history of journalism and technology beginning with Roman missives carved on stone, through the development of the printing press, the telephone and ultimately to Computer Aided Reporting (CAR). Primo suggests that in general technology’s impacts are considered as peripheral influences with the human output still as the central force (Primo and Zago, 2014). Primo proposes that the definition of the end of the human endeavour and the beginning of the technological input has become blurred and technology has so transformed the work of journalism that, to echo the commentary in the previous section of this document, it could be said that technology now also does journalism (Primo and Zago, 2014).

The ubiquity of mobile computing has fundamentally changed the nature of journalism – everyone now carries a mini computer in his or her pocket in the form of a smart phone with an integrated camera, voice recorder, video recorder and web-publishing portal. News is truly on the move (Franklin, 2014). Journalists are liberated now from their desks through the use of technology as Pavlik predicted in his article in 2000 (Pavlik, 2000).

The newsroom has embraced the advantages of technology in increased efficiencies but there are attendant disadvantages of compromised independence and quality (O’Sullivan and Heinonen, 2008).
Alongside the improved speeds of technology, reporters now have wider access to public and governmental sources of data that would previously have been difficult to parse through, leading to the potential of an even stronger role as society’s guardian. Conversely O’ Sullivan intimates that there is still a question outstanding as to whether the new digital mediums empower or compromise the watchdog role (O’Sullivan and Heinonen, 2008).

**History**

Journalism has always relied on technological innovation from the printing press, the typewriter, the telephone and the tape recorder. The Internet and the growth of data signal new means are necessary to process facts and figures.

CAR, Computer Assisted Reporting, was a popular means of using computation to statistically analyse public data and became ubiquitous in journalism, particularly in the US in the 1980s and 1990s (Lewis and Usher, 2014), (Coddington, 2014). This established the technical incursion into newsrooms and its descendant technologies are now the various elements known as computational journalism that will be discussed in detail subsequently. Gynnild goes so far as to say that CAR built the foundations of the current technical infrastructure of newsrooms and normalised the use of digital tools to process quantitative data (Gynnild, 2013). As a result technology and its tools have blended themselves into the day-to-day work of the journalist (Primo and Zago, 2014). Primo also made an interesting point that previously technology had been regarded as an intermediary in the journalistic workflow, a necessary aid passed along the pipeline to support the sourcing and publication of the story. However, with the growth of data, this may need to be reconsidered in the sense that technology is now often the intercessor, and actually can change the flow of composition through the workflow (Primo and Zago, 2014).

Both Primo and Carey caution against overstating the importance of the role that technology plays in a reimagined definition of journalism (Primo and Zago, 2014), (Carey, 2000). Technology should seem as fundamental and natural to the journalist as paint and paintbrushes are to a painter. What is changing is the move to a greater integration of technology and how that can be achieved in a seamless manner. Primo suggests moderating the attention given to the technological impact on ‘co-creation’ in journalism in order to avoid the ‘fetishism of digital artifacts’(sic) and hence a flavour of technological determinism (Primo and Zago, 2014).

Carey for his part insists on the independence of journalism from technology. For him journalism is a ‘peculiar way of using […] technologies rather than the technologies themselves’ (Carey, 2000).

Technology is now an embedded process in creating the news, and Pavlik and McIntosh highlight this convergence of computing, telecommunications and media in the digital platform(Pavlik and McIntosh, 2004). The question now is how much more will technology integrate with journalism?

**Increased use**

Technology has changed the way journalism is done. All media institutions are now expected to have an online presence. With social media the news is out there in a matter of seconds and journalists cannot afford to ignore these sources.

The Internet in the newsroom has moved from the wondrous and different to the mundane and has been an agent for change both in the story telling process and with the sources of those stories (O’Sullivan and Heinonen, 2008).

As with other professions, the information age has altered the industry significantly (Kaul, 2013a). The proliferation of cheap and often free open source tools that are highly portable and powerful
has moved the newsroom to the field (Kaul, 2013a). The standard practices of finding and verifying sources and meeting tight deadlines has been streamlined, with Kaul suggesting that digital journalism could be a strong disruptor to ‘shoe-leather reporting’ (Kaul, 2013a). Kaul does warn that access to such tools and the array of data available could precipitate an increase in plagiarism (Kaul, 2013a). Indeed much of the scepticism about the convergence of technology and journalism relates to ‘shovelware’ or the recycling of news (O’Sullivan and Heinonen, 2008) and what O’Leary has called ‘snowball’ or iterative journalism (O’Leary, 2014).

Kaul further hypothesises that the increased use of digital tools may actually be obscuring the truth in some areas. There is a need to marry the impacts of technical innovation on the media landscape with the core tenets of investigative journalism and principally with the public interest role of journalism. He outlines the fact that journalism has historically been very good at highlighting the issues of the day but for whatever reason, lacking in providing sufficient context around this for a deeper understanding (Kaul, 2013a). The rush to innovation must not overshadow this fundamental purpose of journalism.

Benefits

Spyridou et al contend that the use of technology are empowering journalists to do their jobs better (Spyridou et al., 2013).

On the convergence of technology and journalism, while stating the he does not intend to offer a technologically deterministic view of new media, Pavlik asserts that ‘convergence merely holds the promise of a better, more efficient, more democratic medium for journalism and the public in the twenty-first century’ (Pavlik, 2000).

There are worried voices about how far this technology convergence will go particularly as companies such as Narrative Science have successfully used algorithms to automate the writing process, albeit in a very specific area. While this will be reviewed in the subsequent sections, Roose would contend that the automation of both sourcing and writing could position journalists at the heart of more interesting work (Roose, 2014).

2.3 Computational Journalism

While there is a wide range of literature on the impact of technology on journalism, it is only since 2006 that there has been an attempt to label the relationship between both practices. This attempt resulted in what is described as computational journalism. The phrase has been attributed to Irfan Essa (Coursey, 2013). Flew described it simply as the application of computing to journalism (Flew et al., 2012) while Hamilton and Turner expand the concept further by defining it as the combination of data, analytics and social science practice and they see it as a direct descendent of CAR (Hamilton, 2009).

Computational journalism is an umbrella description of the use of computer science and technology in the journalism sphere. While the phrase suggests the use of analytics and algorithms, it also includes data gathering and storage and areas such as social media and data visualisation techniques. These activities and tools are used in conjunction with, and uphold, journalism’s first principle – ‘the public has a right to know’ (Siegfried, 2007). Moreover, the use of computational journalism has strengthened the other core values of journalism – accuracy, verifiability and independence (Center, 2014).

Aspects of computational journalism that cover data gathering and processing have allowed for easier access to, and contextualisation, of information (Diakopoulos, 2012). Kim Pearson claims that
computational thinking and processes have been influencing journalism long before the current disruption to the industry by the collaborative nature of Web 2.0. She declares that the argument about whether journalists need to be digitally literate is over. Journalism today expects it (Pearson, 2009).

This section reviews the literature on the three main components that come under the umbrella term computational journalism - data driven journalism, automated writing and the use of social media.

2.3.1 Data Driven Journalism

This is a new term that is used to describe the specific processes around data mining and manipulation in journalism. It is ‘big data’ for the newsrooms. The process cannot simply be reduced simply to journalism with data. Its differentiating feature is the combination of the mores of traditional journalism, the “nose for news”, with the scale of the data now accessible to journalists (Gray, 2015). Gray indicates that data can be the source of the story to be told or the supporting tool to a story, or both (Gray, 2015).

Data driven journalism hinges on a degree of democratisation of data - the use of open-source data, sometimes crowd sourced, that would have been previously unwieldy to extrapolate from and translate to visual forms for the reader. It places journalism at the centre of the data maelstrom that has been provoked by the Snowden revelations. De Maeyer et al identified access to public data as one potentially limiting factor in the growth of data journalism – despite public pronouncements on transparency in government, the data is often difficult to work with as it is not digitised or timely and it is often fragmented as it comes from different systems and departments (De Maeyer et al., 2014).

Kenneth Cukier, writing in the Economist in 2010, expounded on the positives of abundant available data but also warned of the impending issues with storage and privacy concerns. He quotes Alex Szalay who claims that the very proliferation and free access of data is making it more inaccessible - it requires a movement from journalist to data scientist and this concept is explored later in this document (Cukier, 2010).

Data is often processed by the use of analytics that can be as simple as an MS Excel formula or as complicated as an intricate statistical calculation, depending on the data type and the skills of the journalist. Kessler has highlighted that analytics can be a boon to storytelling as they can link data and discern patterns more accurately than a human (Kessler, 2012). Visual analytic tools can present the data in an interactive and pictorial format that readers can engage with. Journalists can use analytics to track stories and find sources more efficiently. This can lead to easier and faster access to data and increased data accuracy.

Paul Bradshaw, on his Online Journalism blog, has been to the forefront of documenting the workflow of data driven journalism and his outputs have been used in data driven journalism courses worldwide. The flowchart in Figure 4 below illustrates the many tools and methods of gathering data. The flow may differ from project to project depending on the attributes of the data being sourced.
2.3.2 Automated writing

Currently an area under debate is the increasing occurrence of the use of algorithms to construct whole articles and the publishing of these same articles by credible news outlets such as Forbes, the Wall Street Journal and the New York Times. The company at the heart of this innovation is Narrative Science, formed in Northwestern University USA from a partnership between computer scientists and journalists (Carlson, 2014).

The process uses Artificial Intelligence and marries analytics with natural language study to read data, using data mining techniques, and convert the information into stories in a matter of seconds – the process labelled ‘robotic journalism’ by Latar (Latar, 2015).

There is much online discussion on whether journalists can ever be totally replaced by algorithmic writing (Kessler, 2012), (Shirky, 2009), (Latar, 2015). While the majority of articles are based on numerical facts, such as sport results and financial reporting, there has also been a movement towards using the process to comment on news items such as the reporting of the Los Angeles earthquake in 2014 (Oremus, 2014). Sarah Kessler has also posited that editorials may in the future be automated by algorithms (Kessler, 2012).

Stephen Levy has reviewed many of the articles and concludes that they don’t read like a robot has written them. He interviewed Kristian Hammond of Narrative Science who reassured him that robots
would not replace journalists; rather that the “universe of newswriting will expand dramatically, as computers mine vast troves of data to produce ultracheap, totally readable accounts of events, trends, and developments that no journalist is currently covering” (Levy, 2012).

Michelle Lowery echoes this sentiment – she concludes that the tools are useful but will still need human input (Lowery, 2012).

On the other hand, the use of software to write articles is a new phenomenon (Clerwall, 2014) and for this reason there are fewer studies on the automation of the news generating step of the journalistic workflow.

2.3.3 Social Media

“Social media is the term given to a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0 and allow the creation and exchange of user-generated content” (Kaplan and Haenlein, 2010). It supports the democratisation of knowledge and information, transforming people from content consumers into content producers.

Twitter and YouTube are the principal sources of data but there are many.

The first survey of the use of social media by Irish Journalists at the end of 2014 highlights the increased importance of social media platforms with 99% of respondents using them regularly, Twitter in particular (92%), and with half of all journalists surveyed confirming that they use social media daily to source content (Heravi et al., 2014).

However, in line with other studies and analysis, on the quality of online information Heravi found a high level of distrust about the data (64%) and plotted the reasons in the figure below:
2.4 Impacts of Increased technical automation in Journalism

2.4.1 Automating journalism

In addition to using computation to source and write stories, automation also extends to the use of recommender systems to use social media to create content and the use of tools to access large-scale public information databases. As Gynnild has observed, the exponential increase in the scale and variety of data available currently is impeding rather than enabling the decision making process and quotes McNair in revealing that “Information surplus is identified as a main constituent in the transition from a control paradigm to a chaos paradigm in the globalized news culture” (Gynnild, 2013). Significantly, despite concern about newsroom’s ability to handle large data in a meaningful manner, Flew has identified the use of data and trend identification, particularly in public-interest sources, as an integral part of journalism’s role in a democratic society (Flew et al., 2012). Coddington has written the movement from collecting data to processing involves an increased level of exploratory analysis (Coddington, 2014) that would indicate that data could be a means by which journalists can provide more thoughtful, investigatory content.

The use of statistical reporting through CAR had introduced degrees of computational thinking into the practice of journalism and Kim Pearson has argued that modern news-gathering already encompasses new approaches that are reliant on a wide-ranging toolset that includes programming, data scraping, database skills and visualisation techniques (Pearson, 2009). Pearson does introduce an interesting thought which is that these skills need to be backed up by learning practices that specifically educates journalists on the ethical, transparency and objectivity skills that can integrate computational skills fully into journalism (Pearson, 2009).

There is also some positive commentary around the benefits of using algorithms to write the stories that most journalists dislike writing, potentially freeing them up to create more editorial and investigative pieces (Roose, 2014).

Another positive mentioned by Aeppl in the use of algorithms to write the news is the fact that it is facilitating role change, rather than replacement, in some instances. He uses the example of Credit Suisse, a client of Narrative Science, who before using their service used to use 20 employees, dispersed around the world, to painstakingly summarise 15 companies’s stock activity for future projections. With the new technology Credit Suisse is now able to cover 5000 companies, quality has been standardised and the twenty staff have been able to concentrate on producing detailed report on core companies which is a more interesting task for them (Aeppl, 2014).
2.4.1.1 The augmented workflow

Journalism has seen an increasing impact of technical tools in the workflow process. Computational journalism has enabled the workflow to be refined and augmented by the use of technological tools and aids at different touch-points along the process. Most likely prompted by the increased use of large data stores and the move to create an online content stream, newsrooms now include resources that have defined coding and data visualisation skills (Lewis and Usher, 2014) - what Parasie calls “programmer-journalists” (Parasie, 2012).

The authors would comment that the existing literature is deficient in the respect of the analysis of the workflow coming from a technology point of view. While the detail of Figure 4 above is useful in the context of documenting the new steps that have emerged due to the increased data inputs used to generate stories, the process would benefit from being broken down to its constituent parts into a journalist-computation feedback loop model. This would clearly identify those processes that could be automated or aided by technological tools, and equally those areas where the experience of the journalist adds value at an intrinsic level that cannot be replicated by computers.

2.4.1.2 Creating new Communities of Practice

There is an increasing confluence now between coders and journalists leading to the growth of social movements such as the Hacks/Hackers groups globally (Lewis and Usher, 2014).

However, the exact integration of the two professional threads needs to be quantified before it can be concluded that this is a natural union. The founders of the originating Hacks/Hackers group (Gordon, Herman and Philofer) were wholly from journalism and the inference to draw is that there was recognition of a growing need to engage with developers and explore the possibilities of a closer alignment with development and innovation (Hack Hackers, 2010).

The Dublin chapter of the Hacks/Hackers group was started by some of the leading lights of the computational and data-driven journalism scene in Ireland, and has attracted a mix of journalists and coders, but growth has been slow.

2.4.2 Using computation to strengthen Journalistic Principles

“Society doesn’t need newspapers. What we need is journalism” (Shirky, 2009).

Coddington and Diakopolous both highlight the issues of transparency in computational journalism simply by the nature of the work – algorithms are black boxes that require cost and time to deconstruct for public consumption (Coddington, 2014), (Diakopoulos, 2013). This mirrors the tensions between the ethos of the separate practices of technology and journalism. The authors have investigated this issue in more detail in a later section of the review.

Conversely, publishing the methods by which it has built stories augments data journalism and it is an editorial decision that has served The Guardian newspaper well. Generally it is seen to be adding to the story rather than undermining it (Coddington, 2014).

In a more worrisome trend, computational journalism is seen to be breaking down the ethical barriers between the advertorial and the editorial - online websites for news outlets are proliferated with advertisements, some of them using algorithms to target the reader directly. There are increasing concerns about the ethical decisions being taken on the online sites that would never be sanctioned on the printed newspaper (Pavlik, 2000).
Singer noted that the essential relationship between the journalist and the news, even if this is data-focused, remains the same: “journalists continue to see their primary role as interpreters, rather than mere gatherers and disseminators, of information” (Singer, 1998).

2.4.3 The limits of Automation in Journalism

Many of the barriers to increased automation in journalism come from journalists themselves. Journalists do not see themselves as programmers and see technical innovation as a distinct and separate development within the industry, rather than being a core element of the job itself (Lewis and Usher, 2014). This segregation has been further emphasised by innovative journalists and companies such as Narrative Sciences who are at pains to frequently iterate that the changes they herald are a natural support to existing journalistic practices and not a threat of wholesale change (Coddington, 2014).

Fundamentally, the journalist will always remain at the heart of the news process because of their intrinsic humanness that cannot, as yet, be replicated by machine learning (Steiner, 2012). The ability to form emotional connections with readers, apply cognitive abilities to filter offence, recognize subtleties, be creative and ultimately be accountable ensure that any automated process will still require human eyes at some point in the workflow (Mathews, 2014).

However automation of both sourcing and writing could position journalists at the heart of more interesting work; as Roose has written “rather than putting us out of work, it might free us up to do more of the kinds of work we actually like” (2014). Automated Insights, which along with Narrative Sciences is one of the leading brands in the sphere of automating stories in the newsroom, has responded to criticisms that they are removing the journalist from the news by countering that their software is not seeking to replace journalists, but augment them: “Our technology simulates what a data analyst does. It doesn’t simulate what a reporter does all that well” (Roose, 2014).

From a reading of the literature available on the latest innovations in journalism, the authors contend that the level of innovation seems low but that there is a progression of technology in day-to-day journalism. This is largely a positive development but there are peripheral questions that need further analysis, some of which will be outlined below. Essentially the technical tools are just that and no more. They still need to be operated by someone who is, first and foremost, a journalist. While acknowledging that technology has altered journalism and made things easier, Brand contends that the essence of journalism remains intact (2011). Ultimately the tools serve to “supplement rather than substitute” journalists (Hamilton and Turner, 2009).

2.4.3.1 Algorithmic Accountability

Algorithms have become ubiquitous in digital life – they can decide what product advertisements you see, what content is delivered to you and how recommendations are ranked for you. Diakopoulos calls them the new “power brokers in society” (2013).

By their nature they are lacking in transparency in the full end-to-end methodology. They are often intricate, beyond the ability of those not versed in complex coding techniques. In the case of proprietary software, only the inputs and outputs are visible and the interim workings to produce the output are the ‘black boxes’ of Diakopoulos’s seminal work (2013). They could of course be reverse-engineered but not without committing sizable resources, both monetary and personnel.

By the same dictates as those used in traditional journalism, algorithms used to source or shape the news should adhere to the journalistic principles of transparency and objectivity. They should be subject to an auditing function that would cover their specific use in journalism.
For the reader, more fundamental questions need to be raised according to Mercedes Bunz, such as should publications be required to clearly label if an article is generated by an algorithm or what means are being adopted to control and check the outputs? (Bunz, 2010).

The authors note a paucity of information or analysis into this specific area of computational journalism and posit that codes of practice, which in Europe tend to be voluntary and self-regulated, are out of date with the advancements made in journalism in the past number of years. The journalist risks becoming the custodian of a process they do not understand.

2.4.3.2 Copyright and Legal issues

In the US, Lin Weeks has investigated the impacts of automated journalism vis-à-vis existing media law and has noted that automated journalism has removed culpability from the writer and placed it with the publisher (Weeks, 2014). The authors are aware that European media law does not target individual journalists but rather the publishing outlet but observe that a similar study of the use of algorithms and automation in journalism in relation to EU law would be a critical addition to the canon on this topic.

The authors ask if the current law on copyright is sufficient to deal with the new challenges that may be posed by computational journalism. Much of automated writing reuses existing text, either from a database or gleaned from online searches, and the authors express concern that potential copyright issues have not yet been tested as the technology and its adoption is still in a nascent stage.

Before the convergence of technology and journalism, “cannibalisation” of content was not as easy or as fast as it is now in the digital age (Phillips, 2010). Now, content can be taken and repackaged with no attribution as soon as it is published online without having to wait for the printed page and the subsequent print cycle. This puts pressure on the maintenance of any copyright or exclusivity. Phillips also highlights that bloggers see facts as works in progress where online dialog will confirm any variances - the journalist does not have the same luxury. Phillips also quotes Singer “what truth is to journalists, transparency is to bloggers” (Singer, 2007), cited by (Phillips, 2010). There are potential issues that need to be tested in the area of copyright and defamation.

2.4.4 Journalist versus Data Scientist – the evolving role of the Journalist

“In times of profound change, the learners inherit the earth, while the learned find themselves beautifully equipped to deal with a world that no longer exists.” – Eric Hoffer

With the increased use of technology in the journalism workflow, journalists now require a new skillset to capitalise on the new data and analytic tools available to them. Cindy Royal, in her study of the New York Times news department, highlighted the elementary disconnection between the practices of IT and journalism: “The culture of technology is different than that of journalism. They each carry different ideas about objectivity, transparency, sharing of information and performance” (Royal, 2012) and a union of both behaviours requires a new taxonomy that embraces the merging of the two industries.

The degree to which journalism is getting more technical and digitised has both advantages and disadvantages for journalists - it makes their jobs easier but they must acquire new skills in order to capitalise on this potential (Berghel, 2013). New journalists must emerge from universities with the requisite skills to be able to bridge the gap between journalism and computation.
Education courses are starting to include elements of computational journalism in their courses. Long mentions that The Medill School of Journalism at Northwestern University takes an approach that it is “easier to teach journalism to programmers than programming to journalists” (Long, 2015).

2.5 Computational Journalism in Ireland – the current view

There is a dearth of information on the use of computational thinking and analytics in the Irish context but this has of late been addressed by the extensive work of the Insight Analytics group who are looking at the impact of data across Irish society and whose vision is to “create a healthier, safer, more productive world by empowering a data - by individuals, communities, business and governments” (Insight Centre, 2015). It is a joint initiative across University Colleges Dublin, Galway and Cork and Dublin City University but researchers are based across six of Ireland’s seven universities. One of the most significant outputs thus far in relation to journalism is the comprehensive study undertaken in 2014 to survey Irish journalists on their use of social media as a content source.

Within the industry itself, there has been a core group of individuals who have been formative influences on much of the data-driven activity to date, particularly via the kildarestreet.com and thestory.ie sites and they are both the results of painstaking work by journalists and coders who have created innovative ways to access, parse and create content from public and governmental data, much of it in un-digitised formats.

In terms of terms of computational innovation Storyful, a company set up by journalist Mark Little, is creating innovative products to source news primarily through YouTube and Twitter. Content is verified before being released through a content management system to clients that are as diverse as the Wall Street Journal and the Ellen DeGeneres show. The company is a hybrid of journalistic methods and technical algorithms to find content.

The main Irish broadsheets – the Irish Independent, the Irish Times, the Examiner and the Sunday Business Post – have been slow to embrace data and analytics but with the launch of the Irish Times data blog at the end of 2014 there is an increasing awareness that using data to produce stories can be a successful differentiating factor, as it has in, say, The Guardian newspaper in the UK.

In particular, what emerged from the review of the literature is the lack of an academic body of knowledge on the development of computational journalism in Ireland.

The authors are indebted to John O Sullivan of DCU who in 2005 did the first research on Irish journalists and the move towards the digital platform. Early in 2015 Angela Long wrote an interesting paper on the need for increased coding and computational skills on Irish Journalism courses and the authors are appreciative and influenced by this research. The authors highlight the fact that there is a gap in current research, most likely because the topic is only recently a subject for study.

2.6 Conclusion

The literature review has highlighted that traditional definitions of the work of the journalist may no longer be sufficient to capture the changed reality of an industry where technology has become an important partner.

The confluence of technology and journalism has given rise to a new term – computational journalism – which includes components such as the use of social media as a source of news, data driven journalism where data sources are used to both support and suggest stories, and automated
writing software. It is still in an embryonic stage, but what De Maeyer calls “discourse” is an output in and of itself – it is talking itself into being (De Maeyer et al, 2014).

One aspect of the increased use of technology in the industry has been to productise tools that are solutions to the age-old problem of finding the news. This has led to two new streams of development – the use of algorithms and natural language learning techniques to create software that is able to build articles, and the creation of event detector software to find and monitor news trends on social media platforms.

The journalist needs to revaluate their position in this new reality and one of the supports needed is the inclusion of computational skills in journalism courses. While this is beginning to be the case in the US, there is a lag in similar offerings in the rest of the world. In this initial stage of evolution more innovative journalists are building communities to develop ideas, cross-train and provide a platform for support. This is helping to disperse the skills required, particularly in data driven journalism.

Computational journalism, though it has an established definition, is still not sufficiently known or understood by journalists and this may be a contributing factor in the low level of adoption of technology in the workflow. There is a lack of a systemic approach to how journalism works with technology and vice-versa and in the absence of such there is a certain ‘make-and-do’ approach emerging.

The technology, while a boon to the journalist in terms of efficiencies, must also be examined in terms of the impact on the foundations of good journalistic practice - transparency, independence and the role of the journalist as a watchdog. Technological tools used in the journalism workflow must be subject to the same practices of verification and objectivity, as must the traditional work of the journalist.

There must be more study conducted in relation to some of the potential legal issues that arise when sourcing data online and while this has been examined under US law, there is a gap in relation to European law. Additionally, much of what is known as digital or online journalism is outside the control of regulatory bodies prompting a necessary review in order to keep regulatory frameworks in line with changes in the industry.

Computational journalism is still a new concept in Ireland with data driven journalism until recently in the hands of a small number of practitioners. This has begun to change in the past year but Ireland still lags countries like the UK in the adoption of new technologies and tools.

The literature review, despite the fact that computational journalism is such a recent source for study, revealed a large number of high quality research papers on which to build this study. All the research to date is operating at the initial stage of adoption of computational journalism so it may be a few more years before the full extent of possibility and change in journalism can emerge. Journalism, like many industries, is struggling to adapt as established business models become obsolete - computational journalism may prove to be a solution to some of those problems.
Chapter 3 Methodology

3.1 Introduction

This chapter describes the methodology applied to this study of the extent of automation using components of computational journalism in Ireland. It discusses the philosophical approach, the methods and the type of data analysis employed. The choices in terms of the methods and strategies implemented are discussed and supported by the theoretical arguments that drove them.

The philosophy underlying this study is Pragmatism. Two methods were utilised to answer the research question on the growth of computational journalism in Ireland: interviews and the attendance on an online course on data driven journalism. Based on the utilisation of the two methods, both of which are driven by a qualitative analysis, the approach can be defined as a multi-method research (Saunders et al., 2009). Adopting a pluralist approach allows the researcher to give consideration to different dimensions of the research context including the capabilities and characteristics of the researcher (Mingers, 2001).

Of instrumental guidance to the identification of Pragmatism as the philosophy underpinning this research was Creswell’s seminal work on research (Creswell, 2003) which was one of the suggested readings for the module. Determining and understanding the knowledge claim is a crucial step in the development of the methodology since, as suggested by Creswell, the way a research question is answered is utterly influenced by the philosophy.

Additionally, as outlined by Mingers (Mingers, 2001) and Creswell (Creswell, 2003) it is important that the personal characteristics of the researchers are taken into account when developing the methodology. For example the choice of interviews and qualitative analysis as the primary method of inquiry was determined by the preference of the researchers for interviews as an opportunity for a more in-depth investigation of the subject of study.

Finally, crucial to the development of the methodology was the review of the literature. The authors noted that interviews had been primarily used in many of the studies previously undertaken in this research area. This is further expanded in the literature review section of this chapter.

Saunders et al. (Saunders et al., 2009) provided a practical guide in the development of the framework. Once Pragmatism was identified as the driving philosophical approach, the model they propose (the ‘onion’ model) was used as a step by step guide in building the methodology - starting from the definition of the knowledge claim to specifying how data would be collected and analysed.

3.1.1 Knowledge Claim

After an initial study of research philosophies, Pragmatism was the natural choice. Both researchers identified themselves with its principles and in particular with the way the ‘relationship’ between the researchers and knowledge is depicted in the tradition of pragmatism: “for Pragmatists knowledge exists in the form of statements or theories which are best seen as instruments or tools; coping mechanisms, not once-and-for-all-time truths” (Bryant, 2009).

In accordance with the pragmatic approach, the researchers were more influenced by the need to investigate the topic for research than the methods used to conduct that research. The researchers took advantage of the freedom that a pragmatic approach gives in terms of choices of methods. This resulted, as mentioned above, in the use of two methods (the interviews and the online course). Also, as highlighted by Goldkuhl, Pragmatism is characterised by an inclusion of active and constructive knowledge (Goldkuhl, 2012). This is reflected by the choice of adopting Action Research Methodology
as the secondary method, the aim of which was to gain a better understanding of the industry and in particular the tools and processes available to modern journalists. The researchers felt this would better inform the choice in terms of interviewees and the interaction with them.

As pragmatists, the researchers also recognised that research cannot be isolated from social, historical, political or other contexts. For this reason in the conclusions of this report the authors discuss implications of this research that open the way to new areas of investigation, in a broader context than the one in scope of this research.

3.1.2 The inductive approach

In this research, the relationship of the researcher with the theory is defined by an inductive approach. The research in fact did not start with an explicit theory; on the contrary the findings discussed later in this report are a result of an evolutionary process that began with the review of the literature, which strongly influenced the formation of the research question, through to the data collection and analysis process. At the beginning of the project when the area of research was identified in computational journalism it was not the intention of the researchers to focus on Ireland, as the assumption was that Irish Journalism was not mature from a computational perspective (Long, 2015). In fact a more comprehensive review of the local literature formed the research question and suggested that there was potential to start a discussion on computational journalism in Ireland and with it the first budding of the theoretical conclusions that would fully mature through the interview phase.

As highlighted by Saunders, the authors adhered to the broad view of the inductive approach to ensure that the data gathering from the subjects, specifically chosen for their particular expertise, was also aligned with the aim to contextualize the study within the totality of the wider industry. The predominance of qualitative data used in this research is a consequence of having selected an inductive approach. Studies based on an inductive approach tend to be based on a small sample of subjects and not on a large number which is usually more appropriate where the approach is deductive (Saunders et al., 2009).

3.1.3 Literature review

The initial investigation conducted on this research question consisted of the review of the current literature. Many previous studies had used interviews techniques, which the authors felt would be the best method of gathering quality data.

Royal used on-site interview techniques on her case study of evolving journalistic roles on the New York Times interactive news desk (Royal, 2010). The questions were broad enough in nature to allow detailed responses from which she was able to infer that these ‘programmer journalists’ were well integrated in the news structure. In this instance, Royal assured anonymity for participants which was appropriate given the onsite nature of the observations and the narrowed environment participants came from.

Luoma-aho et al also used interviews, based on qualitative theory, when researching Finnish journalists. The interviews elicited very thoughtful responses and the authors endeavoured to cover as wide a representative sample of employment types within the journalism world (Luoma-aho and et al., 2013).

Karlsen and Stavelin took a similar approach to the researchers of this study in that they identified, by observation of publication and profile, a list of journalists active in the use of coding and other computational tools. They were also constrained by the size of the Norwegian press and newsrooms that had actively embraced computational journalism. Their structured and specific method of
participant selection resulted in eleven interviews. These were more representative of the industry as a whole than the study conducted here, and the researchers were surprised to note the length of the interviews which were between fifty and one hundred and ten minutes. The researchers of this study were influenced by their decision to contextualise the data by linking the answers to the questions. (Karlsen and Stavelin, 2013).

In a broader approach, Deuze et al used survey as a method of reviewing the educational needs of online journalists in Belgium, Germany and Netherlands. The authors undertook due diligence in deciding the definition of a journalist working in an online environment and the researchers of this study noted again the need to carefully analyse the correct participants to engage in studies of computational and digital journalism. The authors also used census reports and unions to size the industries in each country (Deuze et al., 2004).

Closer to home John O Sullivan used both a survey and interviews to research what was the growing movement of journalism into digital medium in 2005. The researchers noted that he was careful to pick as representative a sample as possible and included both national and regional participants. The survey was one he conducted of the main news websites, looking for proof of the adoption of internet features. The author also included a biography of the interviewed participants in the appendices which added useful context in terms of their backgrounds (O'Sullivan, 2005).

These sources all pointed to the worth of interviews as a primary method of data gathering. Without exception, these previous studies had elicited thoughtful and wide-ranging answers to questions, largely because the participants were skilled professionals in the chosen areas. The researchers also noted that in many cases not many interviews were conducted, but rather there was a considered focus on quality and not quantity. Another parallel factor to the approaches adopted by the researchers of this study was the targeted identification of authoritative voices – this was a formalised approach that highlighted a necessity for experienced participants, particularly where the sample size was small.

The researchers of this study were thankful to the work that had already taken place prior to this study and reassured that much of the methodology was similar and had previously been successful in gathering data of high value.

### 3.2 The methods of collection of primary data

This section outlines the methods or strategies (Saunders et al., 2009) used to collect primary data. The primary method consisted of semi-structured interviews and the secondary method consisted of the attendance on an online course in data driven journalism.

#### 3.2.1 Interviews

Interviews were selected as the primary method for a number or reasons. As seen above the literature review unveiled many examples of studies where interviews had been used as a successful method to gather information. Ultimately, interviews were chosen as they would give the researchers the opportunity for a deep (albeit narrow) dive in the area of computational journalism, with access to a great level of detail from subjects chosen as pioneers in the field.

The authors made a decision not to conduct surveys for this research because from the initial reading of the literature it was clear that there was a lack of clarity around what is meant by the term computational journalism in Ireland. This was confirmed by some discussion with the interviewees. The researchers felt that a survey might not elicit sufficiently relevant data in that context.
3.2.1.1 Participants selection

The selection of the interviewees was of crucial importance to the quality of the study. The time constraint, typical of research projects of similar size, as well as the number of researchers, determined a necessary limitation in the number of interviews, which was set at six. The risk with a low number of participants could have been that poor quality information or inadequate information collected from even one participant would have potentially impacted the overall quality of the study. For this reason it was imperative that the research sample would be made up of uniquely positioned and highly qualified subjects that would bring their knowledge and experience to the research. Once the research question was defined the researchers set to identify, based on predefined criteria, the best positioned individuals in Ireland to contribute to build a clear picture of the extent to which Irish journalism has embraced technology. The choice of the participants was inspired by the idea of Communities of Practices – an idea rooted in Pragmatism. All the participants in fact are linked to each other, generally by virtue of shared work experiences, as well as a common interest in technology as a mean of empowering journalists.

The selection of the interviewees was also determined by the fact that the researchers wanted to explore different angles of enquiry. This resulted in a spectrum of profiles that included six distinct positions: the journalist, the data scientist, the innovator, the technologist, the regulator and a barrister expert in digital and online law. In support of these choices a short profile of each of the interviewee has been included in the Appendix.

The participants were originally contacted by email. The email contained a brief description of the project. It was of great gratification that all the invited participants agreed with enthusiasm to be interviewed or in one case a more appropriate contact was suggested, which turned out to be a very positive addition to the research.

3.2.1.2 Details of the interview format

The interviews were semi structured. The primary reason for opting for a semi-structured format was to be able to adapt the questions to the different interviewees, to enable the researchers to follow paths that hadn’t been considered prior to the interview and to follow the leads given by the interviewee. Questions were drafted, scored for relevance and prioritised accordingly. Core questions were selected for each interview and identified as ‘must-ask’ questions. The non-core questions were slightly modified during the interview to allow the interviewer to follow leads given by the interviewees. A sample matrix of the questions from one interview is included in the Appendix.

An Interview Protocol was produced and followed through during each interview. This included a high level discussion of the research proposal, a walk-through of a detailed Information Sheet which required sign-off by the researchers and the subjects, and the Consent Form which confirmed voluntary participation, the consent to recording the interview and the use of the subjects’ identities in the final report. In addition the researchers provided their contact details for further queries.

Interviews are very time consuming and for this reason it was imperative to use the time as effectively as possible through careful planning. Each interview including the arrangements, the travelling times, the physical time for the interview, the transcription time and the analysis, took on average of ten hours each. Accurate planning was crucial to ensure this time would not inflate. As a lessons learned output from the first interview, the interview protocol and the consent form were sent in advance to give the option to the interviewee to review them in advance of the meeting – this resulted in saving time in the actual meeting. The interview duration was also kept to the agreed limits as the researchers were adamant not to take any more time than what had been agreed with the participants. Bell recommends that agreed duration times are respected to ensure that
participants that otherwise may feel their participation takes too long may be discouraged to take part in future researches (Bell, 2005).

Bell’s work was a practical guide in the design and execution of the interviews. This helped in preparing for the potential shortcomings of interviews, for example the risk of bias. While the risk of bias is high, given that researchers after all are only human, being aware of it helped the researchers to strive for objectivity and an open mind.

All interviews were recorded using 2 devices, a primary and a backup. The researchers both attended all interviews and took turns in the roles of interviewers and note taker. The notes underpinned early codification of the data. Interviews were then fully transcribed. The online transcription software called ‘Transcribe’ was used. It was calculated that this reduced the transcription time by 40% - this measurement was based on comparing the times required to complete previous transcription work. The transcriptions were transferred to MS Word and each was stored in a separate document, with the information on the date and details on who conducted the interviews.

3.2.1.3 Lessons learned

After each interview lessons learned were recorded. The following is a list of the main learning outcomes from the interviews, which helped to inform subsequent interviews:

- Send the information sheet and consent form in advance so that the interviewees are not overwhelmed with that detail at the start of the interview. This will also shorten the time of the interview
- If the subject is nervous or has never interviewed before or is unsure with what might be asked, send a specific interview protocol to them in advance with the questions included
- A minimum of two researchers must attend interviews. This allows for an interviewer and a note-taker – it is difficult to be both. Additionally, this allow for two different perspectives
- Semi-structured interviews worked well – the prepared questions were the main theme but there was great opportunity to tease out other opinions in the absence of a rigid structure.
- Use two recording devices, if recording is permitted – one as backup. Ensure it is placed right in the middle of the table and do an initial test that it is recording correctly. Have it on ‘do not disturb’ mode to avoid calls interrupting
- Pick the interview location with due consideration. Background noise can be a major issue and can significantly slow down the transcription
- Keep rigidly to the allotted time
- Compile lessons learned after each interview – did reading the questions from a laptop work better than paper for example? This allows for greater efficiency for subsequent interviews

3.2.2 Online course

The secondary method, prompted by the initial lit review, consisted of an online course in Data Driven Journalism given by the Canvas Network (http://datajournalismcourse.net/course.php). The researchers felt that the course aligned closely with the subjects that were going to be interviewed and would enhance their knowledge of the journalist technology toolkit. The course was a five week online course, and the researchers fast tracked that to do it in a week due to an impending closing date.

This research method can be classified as Action Research, a strategy of inquiry often found in Pragmatism. As the name suggests, in Action Research, the research is in the action rather than being research about action Coghlan and Brannick (Coghlan and Brannick, 2005) cited by Saunders et
al (Saunders et al., 2009). Meyer, as cited by Koshy adds to the description of Action Research defining it as the “ability to empower practitioners, by getting them to engage with research and the subsequent development or implementation activities” (Koshy et al., 2011).

Due to the time constraints, it was decided that only one researcher would attend the course. During the course, the researcher took notes, which were elaborated and shared with the other researcher.

### 3.3 Qualitative analysis

As outlined by Goldkuhl (Goldkuhl, 2012) it may be difficult to reduce complex social and technical phenomena in the IS field to quantitative figures. This contributed to the decision to qualitatively analyse the data collected through the interviews.

Some quantitative secondary data has also been incorporated in the research to augment the information that would help answering the question.

#### 3.3.1 Data Analysis

The methodology used to qualitatively analyse data is a fairly common one. The publication “A general approach for analysing qualitative evaluation data” (Thomas, 2006) served as a useful guide.

The interviews transcriptions were read and annotated by each researcher. The annotations were then used to develop labels to attach to ideas, facts, theories and anything that the researcher found relevant or even surprising when reviewing the content. The codes were then grouped based on commonalities and links. The emerging groups were then categorised into themes. The findings were then compared by the two researchers and in some cases new groups were identified as a result of the comparison, other labels or codes were either merged or disregarded. The result of the codifying process has been elaborated in the following chapter.

### 3.4 Secondary data

Secondary data has also been incorporated in the research to augment the information that would help answering the question. The main source of secondary data was the Central Statistics Office 2011 census results.

### 3.5 Tools

While not an exhaustive list, the following are the main tools and aids used by the researchers:

**Excel:**

- to score participants;
- to build the question matrix and score them based on importance
- to codify and analyse the data
- to build pivot tables and graphs

**Audacity:**

- to slow down the recording (only used initially)

**Transcribe:**
• to transcribe the interviews

Voice recording:

• Android and OS

Endnote

• To manage sources, references and bibliography

3.6 Limitations and constraints of this research

As mentioned earlier in this chapter, time has been the biggest constraint in this research project. For this reason this research is to be intended as a preliminary study of the current confluence of journalism and technology in Ireland and as such it can be defined as a cross-sectional study (Saunders et al.). The time constraints impacted on the decision of the number of methods to use. When the research question was originally conceptualised it was decided that an experiment could be carried out to investigate the perception of the public in terms of automation in journalism.

The experiment would have consisted of publishing 2 articles, one of which would have been written using automated writing software, while the other one by a human; readers of the online newspaper would have been asked to indicate if they perceived the article by either having been written by an algorithm or by a human. This experiment would have been the replication of a similar experiment carried out by Clerwall (Clerwall, 2014).

The number of resources was a secondary major constraint which influenced the number of interview participants.

3.7 Conclusion

The methodology described above proved successful in achieving the type of information that was felt would positively determine the level of quality pursued in this research project. Specifically, the interviews gave a great insight in terms of the confluence of technology and journalism in Ireland. This was further enhanced by the choice of the targeted subjects.

Understanding the theoretical implications at the basis of a research project and the process of building the methodology proved to be a stimulating learning experience.

Having a clear idea of how data was going to be collected and what the motivations were behind the choices made the analysis process, discussed in the following chapter, more straightforward and meaningful.
Chapter 4  Findings and Analysis

4.1  Introduction

The previous chapter described how Pragmatism determined the methodology framework that drove the research and how the inductive approach influenced the qualitative data analysis chosen to answer the research question.

This chapter focuses on the details of the data collection, how data was analysed and ultimately it provides a description of the findings in an attempt to answer the question: is Irish journalism becoming increasing automated and if yes, is there a limit to that automation?

The research findings are organised by themes which have been determined by the analysis of the interview data. The themes are: Changes in journalism in Ireland; Automation; Computational Journalism; Education and Roles; Technology; Regulatory Framework; Potential Legal Issues.

An important premise to make before proceeding with the description of the analysis and the resulting findings is that the literature review influenced the questions, which in turn generated answers that once analysed guided the researchers through the identification of themes. Each of the findings are considered with reference to the literature review. Sub-themes also emerged through the analysis and where they are peripheral to the core research they are discussed in the conclusion chapter of this report. The sub-themes were instrumental in defining areas of future research.

As noted earlier in the report, the primary method of enquiry consisted of semi-structured interviews which involved four key participants and two other participants who were selected based on their area of expertise for supporting evidence. The selection of the participants, and the quality of insight they have brought to the research, is further discussed in the Participants section below.

4.2  How was data analysed?

Each interview transcript was reviewed iteratively by the researchers. In order to identify the themes a three step process was then completed. The three steps were: annotation, grouping and categorisation into themes.

After the initial reading, the transcript was read again, this time with the intent to highlight and annotate any concepts, quotes, ideas and words that were of support to the researchers in their attempt to answer the research questions.

All the annotations were subsequently grouped and labelled according to the area of interest and other common attributes.

The third and final step consisted in the categorisation of groups under themes.

An MS Excel spreadsheet was used for the analysis. The Excel file was created by the researchers following the guidelines of traditional qualitative data analysis techniques. All annotated text was copied into the Excel file. Each text segment or note was given a unique identifier. All notes could be filtered by source name, annotation, group and theme.

4.3  Participants

The interviewees were categorised into two groups - key participants and supporting participants. The key participants were identified because of their position at the forefront of computational journalism in Ireland. The supporting participants were chosen because of their expertise in themes that abutted the research topic.
1. Key study participants

These were the participants that were identified because they were working in varied roles using data gathering tools and techniques in the industry. The interviews of this group were approached in a relatively heterogeneous manner – question matrices were modified to reflect the particular niche that the participant was selected for but there were core questions that were asked to trace common points of interest.

The participants in this group are as listed below and, in an effort to maintain a degree of flow in the remainder of the chapter they have been labelled with the identifier code in the ID column.

<table>
<thead>
<tr>
<th>Name</th>
<th>Title</th>
<th>Selection Rationale</th>
<th>ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gavin Sheridan</td>
<td>Innovator</td>
<td>Blogger, data-driven journalism early adopter, and innovator in the area of product development for journalism.</td>
<td>J1</td>
</tr>
<tr>
<td>Paul Watson</td>
<td>Technologist</td>
<td>CTO of Storyful</td>
<td>T1</td>
</tr>
<tr>
<td>Bahareh Heravi</td>
<td>Data Scientist</td>
<td>Insight Analytics Lead and Irish Times lead data scientist</td>
<td>T2</td>
</tr>
<tr>
<td>Pamela Duncan</td>
<td>Journalist</td>
<td>data journalist in the Irish Times</td>
<td>J2</td>
</tr>
</tbody>
</table>

Table 1 Key Participants

These participants were profiled further to get a sense of how they graphed on a technology versus journalism divide in order to understand any natural biases that might be present in the answers.

The following questions were asked for profiling purposes:

1. How long have you worked in journalism?
2. What is your education/ training background?
3. Do you consider yourself IT literate or comfortable with technology?
4. What percentage of your job is solely based on using technology?

These questions were asked at the beginning of the interviews of the key participants. In addition to helping build up a profile of the participant’s areas of expertise and their split between technology and journalism practices, they helped both interviewer and interviewee settle into the interview and established rapport.

The answers to these questions were analysed to identify how long each participant had worked in journalism, or technology or a hybrid of both. As further supporting data, the participants LinkedIn profiles were used to fill any gaps - three of the participants had a mature and up to date profile, one participant had no profile at all but gave a very detailed chronological answer to the first question.

This allowed the researchers to identify periods where there was an overlap between journalism and technology, which is identified as the Hybrid series in the chart below.
The researchers suggest that, from the graph above and the data gathered from the participants themselves, journalists can move into a hybrid role with minimal training using everyday tools and an on-the-job approach to practice.

Technologists can move into aspects of journalism that utilise their skills but are more likely to be in a hybrid role as the particular skills specific to journalism – shorthand, interview techniques and communication and writing ability – take more time to acquire. This would align with the comment from Emily Bell on journalism unicorns: “There is something about not just being able to think and act like a programmer but also to be able to think and act like a journalist, which is quite demanding” (Lafrance, 2013).

Over the course of the interviews the true sense of a community that has been established among these individuals was revealed in line with the community of practice theory that had prompted their selection. They had been strong influencers on each other and they seem to be representative of a small community of committed individuals that are acting as a stimulus for the slow growth of, particularly, data journalism in Ireland. This shows parallels with De Maeyer et al’s study in Belgium (De Maeyer et al., 2014).

2. Supporting study participants

The supporting participants were identified from the survey of the literature that prompted questions that were not central to the interview structure of the key participants.

They were:

1. Peter Feeney – the Press Ombudsman of Ireland who described his role as: “The job of the Press Ombudsman is to encourage freedom of expression, freedom of speech, and to provide a service for the public who feel that somebody has been [...] is unacceptable that has been published, so you’ve got to kind of get that balance so there’s a lot of work to be done.”
II. Patricia Sheehy Skeffington – a barrister with expertise in digital law and data protection and adjunct lecturer on the information systems course in Trinity College Dublin.

The questions that they were asked were targeted to their area of expertise.

They were not asked any profiling questions – they were invited to be part of the research to focus particularly on the areas of legal issues such as copyright infringement and data protection violations and the issue of regulation of computational journalism. These topics had either come up in the literature review and/or become an increasingly heightened area due to the key participants’ answers, or were areas that the researchers felt deserved more focus.

4.4 The changes in Journalism in Ireland

The discussion on the impact of automation on Irish journalism cannot be isolated from a more general discussion on Irish journalism itself. In this context, the interviews aimed at understanding if what had been discovered as part of the review of the literature, with regard to the changing nature of journalism, is true for Ireland too.

The definition of journalism was discussed and responses confirmed that journalism in Ireland is undergoing a transformation process based on which traditional definitions of journalism may no longer apply. The influence of blogging on the Irish scene is an example of the evolving shape of journalism.

One of the interviewees defined journalism as “an artificial concept in some ways about how you manage information and what you do with it”.

Irish journalism has been impacted by technology and as a consequence it is changing and it is no longer happening exclusively in newsrooms.

Blogging for example became an essential component of the journalism taxonomy. One of the participants noted how, as recently as 2009, bloggers tended to be ‘outsiders’, removed from the traditional journalist background. Since then, ‘traditional’ journalists have embraced blogging and other forms of social media, adopting them as an integral part of their work.

In fact as shown by the 2014 study carried out by the Insight Analytic centre, Irish journalists are among the heaviest users of Social Media in the world (Heravi et al., 2014).

Storyful represents a further example of how Irish journalism is undergoing change and how for Irish journalists the newsroom is not necessarily the main career choice. The company, which operates globally and who was defined by its CTO (one of the participants) as a hybrid between journalism and technology, employs between 40-50 journalist and about 17 technical staff, including developers, designers and product owners.

Finding – The article is no longer the only output of the journalistic workflow

In the general discussion of how Irish journalism has changed as a consequence of the new technologies, an interesting finding was an emerging view of how an article is no longer the only output of the journalistic workflow.

This view was strongly represented by the more technological among the two journalists interviewed. Coming from a technical background seemed to allow the journalist to consider information as it is seen in an Information System context, more than a journalistic context. Therefore information becomes relevant not only to generate news but also as a potential product to be monetised.
The monetisation of information came up as a recurrent theme which, given the scope of this research, has not been expanded further but that nonetheless has confirmed what had been mentioned in the literature review in terms of how digitalisation of journalism is indeed opening up the prospect of the monetisation of information.

All the participants agreed that technology is changing the way journalism is happening in Ireland even if the change is perceived as slow. There is an acceptance that “knowledge management and Information management in the news rooms is basically non-existent”.

An interesting remark in terms of the confluence of technology and journalism was the comparison between the profession of journalist and the work of a programmer. One of the participants has close ties to developers and has noticed that there is a narrowing of the gap between the two professions.

4.4.1 Future of journalism

The interviews prompted questions in terms of the future of journalism and how it will evolve, at an Irish as well as a global level. The question: “Are you worried about the future of Journalism?” revealed a dual sentiment in this respect.

On one side there is a concern in terms of the certainty into the future and in particular in terms of future revenue streams. The ultimate concern is that extreme automation could impact on employability. This point is further discussed in the next section of this chapter.

On the other side, despite these concerns, there is a certain enthusiasm for the prospects of journalism and data journalism. One participant only recently involved in data journalism commented:

“I’m really, really excited about the prospects of journalism and especially data journalism. It’s probably the most exciting time in my career in that I have this niche that I’m developing and learning all the time and I really am enjoying it and I do see a future for it”.

The researchers would argue that this particular view requires a bigger scale research and sample to be validated as an expression of the general sentiment across journalists. The small sample of participants and the narrow focus in terms of their involvement in existing data journalism activities might indicate a degree of bias that a larger and more representative industry sample might not have. In general, the participants observed that the industry was slow to change and this causes frustration among more innovative journalists. Despite this one of the interviewees noted that the industry in Ireland is ready to embrace technology and observed that a sense of willingness is there. There was agreement that technology will have a positive influence on Irish newsrooms and other types of news organisations.

4.5 Automation

Automation represented one of the core themes of this research. There was agreement among the key participants that there is scope for Irish journalism to automate more processes.

The level of automation in Irish journalism is perceived as very low. This echoes news organisations outside of Ireland that, though supposedly more innovative, are still using ad hoc processes. The examples given were two of the most authoritative news organisations, the Guardian and the New York Times. A participant commented that software brings a coherence to how data is used that is probably missing from most journalism organisations.
What emerged from the interviews is that the overall approach to the gathering, parsing and distributing of the news is not systematic and is often made by ad hoc processes or systems that would have the potential to become products in their own right, if developed.

**Finding – Capacity for once-off solutions to become industry products**

This idea of transforming the solutions into products was expressed by one of the participants in particular and is reflected by the way Storyful operates – the company commercialises products developed internally to resolve problems related to the journalism workflow.

Though one of the initial focuses of the research had been the automated creation of the news (particularly by means of automated writing), it was found that most of the product development in Ireland has been concerned with the news gathering step.

According to one participant in particular, this step of the workflow is a necessary area of focus considering the number of events that happen every minute around the world and that news organisations need to discover in the shortest possible time. Event Detection systems have been a particular focus of Storyful which in 2012 developed a system which detected events as they occurred, giving the company a head start before major news organisations who employed thousands of staff knew the events had even taken place. The system would alert the journalist to check a particular event.

The discovery step of the workflow is the most important according to another one of the participants as it can bring real value to the journalistic workflow. Another participant suggested that using systems such as event detectors or RSS feeds would free up time spent by journalists manually monitoring websites every day.

What also emerged was that good opportunities and benefits in terms of automation can come from recommender systems. These systems would detect links between information pieces and recommend them to the journalist, who otherwise would have to manually identify these links. These benefits echo the studies by Gynnild (Gynnild, 2013) and Flew, Spurgeon et al (Flew et al., 2012).

Another area identified for opportunities in terms of automation and systematic approach were the requests of information to public agencies, regulated under the Freedom of Information Act 2014, and commonly referred to as FOIs. According to one of the participants, when requesting FOIs the tendency is for journalists to look at documents and not at databases where other stories may be found. In addition, once the data is collected it is used for a story but not appropriately archived. This results in missing future opportunities to re-use that data for other stories or to find other stories in the data.

**Digital archiving** emerged as an important sub-theme and, while outside the scope of this research, it has been discussed further in the conclusion chapter of this report.

A systematic approach resulting in building novel workflows would help journalists free up some of their time for more challenging and interesting work as suggested by Roose (Roose, 2014). But generally the approach in Irish journalism is not systematic, for example in the news gathering step. Setting up a systematic approach would no doubt require more skills and could initially require an investment in time but in the long run it would result in a faster and smarter approach.
An interesting answer in terms of automation came from J2, who said they would not consider their work as automated, despite it consisting mainly of data-driven journalism activities. This aligns with Cukier’s opinion that the proliferation of data can actually create more work (Cukier, 2010).

What emerged from the interviews was that automation can support the journalist by taking care of most mundane and repetitive tasks leaving them free to focus on the tasks that require his or her skillset.

Finally, automation was perceived by all key participants as a ‘friend’ of the modern journalist as it empowers her or him to do their work. One of the participants added that automation of certain processes is good as far as it doesn’t attempt to replace the human journalist.

J1 was not concerned at all with the risks of automation. Two key participants, who would tend more towards the technology end of the spectrum, did say that while the ambition is to develop sophisticated systems that can support the journalist in the workflow, they cannot foresee the human being wholly removed from the process. In this regard the next section discusses the limits of innovations as they have emerged through the discussions with the key participants.

4.5.1 Limits of automation

The view shared by the participants is that while machines should be given as much work as possible, it is mainly to support humans: “because humans have unique skills that machines can never replicate so you want humans to focus on those problems. Copying and pasting is not a human skill”.

There are certain human skills that machines can never replicate. The technical participants were not confident in particular that artificial intelligence could evolve to that extent. Not in the next 10 years at least, as one of the participants noted:

“we always go by the 80/20 rule here when we build systems... We can build our code to do 80% of the job for you and the final 20% is the hardest, the longest, it’s going to take A.I., it’s been tomorrow for the last 10 years and it’s probably going to be that way for quite a bit longer. Not saying it’s not coming, it will come but it’ll be much more specific, it’ll be less of the general AI. It’ll be much more niche AI. And it’ll be very useful but again it won’t replace a human from gut instinct and that kind of thing”

A specific area that requires human skills is verification. Verification tasks consist of the review of sources and material, often user generated, for authenticity and trustworthiness. This is an important step at the news gathering stage and the interviews established that it cannot be automated, at least not fully.

In qualifying the limits of automating judgement tasks, the participants zoned in on the instinctual gifts that humans possess for element linking through memory recall.

“The human brain is very good at recognising wheat for the chaff, it’s very good at recognising insight. It’s a bit like data scientists work, when you work with large data sets...sometimes the machines wouldn’t be able to figure out a certain amount of stuff, sometimes you need a human to just go I am able to associate that data set with something I know from my experience that I have collected over my lifetime... So that’s quite hard for machines to replicate. But my belief is that you want to push the boundaries as far as you can and automate as much of it as possible. And I think that’s a good thing.”

As summarised by one of the participants, human judgement cannot be completely replicated.
Automation does raise questions about the future role of the journalist. The participants, in the majority, were certain that the human could not be removed from the workflow so at least cognitively a self-imposed limit to automation emerges - it cannot be conceived of. Even a very innovative company (Storyful) doesn’t aim at replacing journalists.

If journalists fear the competition of automation, can they fully embrace it? The researchers believe that this results in a challenge that the technologist and the journalist have to resolve collaboratively.

The literature review highlighted that one of the main barriers to automation comes from journalists themselves. Technical innovation is distinct from the core workflow and, as many of them are not versed in technology, they do not see the capacity for improving their jobs (Lewis and Usher, 2014).

4.6 Computational Journalism

In the discussion on automation in the previous section specific focus was given to the concept of Computational Journalism and consequently its main components: data driven journalism, automated writing and social media. In the following subsections the main findings under these themes are discussed.

The terms Computational Journalism was first introduced to the participants as part of the invitation email that was sent to each of them. The initial reaction to the term by some of the participants indicated what was then confirmed during the interviews, namely that not everyone was familiar with the concept. Two of the participants (one with a technical profile and the journalist) were hesitant to be interviewed as they believed that they did not have any knowledge to impart.

A third participant admitted having to research the term before meeting for the interview. Two of the remaining interviewees were familiar with the term but associated it with data driven journalism only. This uncertainty with the term confirmed what emerged from the review of the literature. Computational journalism is a very recent umbrella term that does not have a set-in-stone definition. While universities like Columbia do have courses on computational journalism, in Ireland the concept is not well known.

During the course of the interviews, a natural understanding was uncovered by one of the participants that best describes the term - technology filters information so technology and journalism have to meet somewhere in the middle.

All participants agreed that technology and journalism are now seen as symbiotic. Compared to a few years ago the shift to technology has been dramatic. It was acknowledged that there is a confluence of technology and journalism in Ireland though it was still embryonic.

4.6.1 Data Driven Journalism

Data Driven Journalism has emerged as the most dominant theme of this analysis.

**Finding: data driven journalism is growing in news organisation in Ireland and data is now a fundamental part of journalism despite the low numbers of practitioners.**

Journalists see in data opportunities to create stories validating Gray (Gray, 2015). However, the attitude in this respect was different between the data journalist, who would tend to look at the data as a consequence of an idea and the attitude of the technologist journalist, who looks at data not only as a consequence of an idea but also to find ideas in the data.

The source of the data was surprising to the researchers, as social media data was not used as a primary source to generate stories. This contradicts the literature review and the percentages of
journalists surveyed that would use Social Media as a source for their work as part of Heravi’s digital
survey (Heravi et al., 2014). The researchers suggest this is because the interviewees in the main
were focused on getting mainly public-interest data.

A consequence of the increased use of data in journalism is the corresponding increase in the
workload to process it. There is so much information that it is difficult to know where to start. While
there are more and more opportunities for data journalism there’s also a data glut and it can be
difficult to navigate through.

This issue in particular was brought up by one of the interviewees who said: “Accuracy depends on
the quality of the data being put out and on the individual efforts of the journalists that are looking
at it and treating it and I think journalists have to be responsible in their treatment of the data and
the checks that they go through in order to ensure accuracy.”

This opinion reiterates De Maeyer’s commentary on the quality and availability of public data as a
limiting factor – it is either unusable or needs a considerable effort to render it serviceable (De
Maeyer et al., 2014).

The mentality of the digital native is to put everything online. This is resulting in the building of a
personal digital archive and the implication is that this must be mirrored at governmental and public
body level.

**Finding – Data Driven Journalism engages the public in Watchdog journalism**

It was revealed that data empowers the reader to get involved and act as the watchdog. With data it
is easier for the public to challenge the piece of news. Data based stories are usually accompanied
by the methodology and the actual data in support of the story. Some readers find this stimulating
and participate in the discussion occasionally by challenging the interpretation of data. This has been
the experience of the Irish Times Data blog.

The participants felt that the more data there is out there, the fewer places there are to hide and
therefore it is good for democracy. Ththestory.ie and Kildarestreet.com and their founders (one of
whom was interviewed) are the best examples of data driven advocacy in Ireland and have been an
influencing factor on the participants.

This researchers feel this goes some way to answering the question posed in the literature by O’
Sullivan and Heinonen as to whether digital mediums empower or compromise the watchdog role
(O’Sullivan and Heinonen, 2008). From a democratic standpoint it empowers the public to take a
more engaged and informed part in debate but the data sources themselves raise other issues about
regulation that are discussed later in the chapter.

The visual communication of the stories generated by data driven journalism captures the interest of
those readers less inclined to spend the time reading long texts, though the interviewees did note
that use of graphical interfaces are seen as ‘shallow news’ by more traditional journalists in the
newsroom.

As a counter-weight to this, the data journalist interviewed was convinced that data driven
journalism and traditional journalism work symbiotically together – one inspires the other. She
found that as she finishes an element of a data story the traditional element kicks in and goes to get
the face to face interviews to corroborate the story out of the data. This viewpoint was a contrast to
Kaul’s theory that digital journalism could be a strong disruptor to ‘shoe-leather reporting’ (Kaul,
2013b).
The interviewed journalists agreed that the toolset of data driven journalism can be basic and MS Excel emerged as the primary tool. There is sense that with the technology getting better interpreting data is becoming easier.

However there is an acknowledgement that the journalist that can do data analysis has a new independence and no longer relies on someone else to get the data.

Data driven journalism is becoming an accepted element in the news cycle and the question now is how prolific will it become.

4.6.2 Automated writing

One of the areas in automation investigated by the researchers, was ‘automated writing’.

The questions asked were:

- Do you think that the way the news is told will be different with the increased use of algorithms and data analytics?
- In your opinion what is the current level of adoption of automated journalism in Ireland?

Key participants were asked if they were aware of any software-generated articles published by Irish news organisations. The participants answered that they were not aware but they did not believe such software, if used at all, would be heavily used in Ireland.

One of the participants had conducted previous market research on the technology used across the journalism industry and noted that his research did not uncover the usage of such software.

In terms of the implementation of such software in general, the views were different. One participant for example said that the best examples they had seen are still quite poor and they are usually limited to news on earthquakes (directly generated by earthquake centres) and company file reports. The participant added that such software is not sophisticated enough.

Another view was that such software could and does serve well in areas like sport reports and can contribute to resolving the issue of freeing the human fromformulaic writing that is best suited to algorithmic writing.

Business journalism was considered to have great potential for automation. Current processes for analysing company records are mainly ad hoc and manual. There are platforms that would support the automation, or tools like RSS feeds but only the most innovative among journalists use them.

J2 answered that she had never used such software and that using it would be “out of question”. She felt that letting software write a story to which you tag your by-line would be like getting someone else to write the story for you. The journalist suggested that having sophisticated software to fully automate the transcribing, would be much more beneficial to the journalist than an automated writing software.

The area of automated writing did not emerge as an area of concern among the participants and one added they didn’t believe it would result in full automation of writing.

Finally, the most fundamental issue with automated writing seems to be that it is not really answering a fundamental problem about news discovery or news distribution, which remains the main focus for journalists.
4.7 Education and Roles

A repeated theme to emerge from the literature review was the need to revaluate the definition of the work of a journalist due to the increasing convergence of technology in the workflow. Technology is impacting how journalists work and there is an increased need to realign the educational courses producing journalists in Ireland with the emerging industry needs. Allied to this is a need for journalists themselves to identify with the evolving role of the journalist.

These two themes were core to this research and analysis of existing studies and in addition to the answers provided by the participants, the researchers looked at the amount of courses available internationally to see how far computational journalism components have been integrated into journalism education.

4.7.1 Education

The key participants group were an interesting mix of journalists and technologists as illustrated by the chart in the Participants section above.

In terms of their professional competencies, many of the data and technical skills were largely self-taught. One participant had completed a specific third level qualification in journalism and data skills were gained from a short three-day course in the UK.

Specifically on the issue of education a question was asked the three participants working in journalism: “Do you think that journalists and journalists’ educators are equipped with the right tool set and skills to capitalise on the increased use of technology?”

All three brought different strengths and biases to this question but all answered strongly on the need for education providers to embrace the change, particularly towards data-driven journalism. There was agreement also on the immediacy of this need for change.

Academia has not identified the growth of data in journalism as a source for new courses or modules.

While it is natural that academia will trail industry due to the more dynamic cycle of change in the latter, the speed of change in journalism in the past five years in particular is unprecedented. They admitted that even as practitioners in the field of data journalism, it was a struggle to keep up with new developments. Education providers would struggle even more.

One participant highlighted that in Irish universities there is very little focus on the new data mining techniques and using data to drive out stories. The researchers suggest that this is still a nascent area for educators to get familiar with and believe the most important comment was from a participant who suggested that as long as courses still use established journalists, which by default will tend to mean journalists of longer tenure who may not be catching the digital wave, the status quo will be maintained. But another respondent emphasised that the need is great now, not tomorrow.

The press ombudsman also spoke on this topic and his perspective was that a journalist well trained with using digital journalism would be better equipped to deal with the legal and regulatory issues that might arise. He observed that:

“it all has to be part of training journalists today and the whole IT part that is, that no journalist should come out of any college without being absolutely on top of it, in terms of IT. Because, that is the world they will live in, the world they will earn their living, and also the world in which they will gain most of their information.”
The responses broadly agree with the literature on the topic though they do focus on the need to include new computational theory as soon as possible. Long’s recent paper has stated that educators do face a difficult challenge to balance the need to teach fundamental journalism theory and new programming skills. She takes a practical approach that at the very least, even if journalists don’t learn sophisticated coding, that they should be able to communicate with experienced coders and work with them (Long, 2015).

The researchers spoke to the participants about their perceptions of Irish journalism courses when juxtaposed with international offerings. One participant has been involved in training initiatives in Madrid. In an Irish context, there is beginning to be a move to insert data journalism modules into existing courses but there are no plans that anyone was aware of to introduce degrees on the subject. There is a belief that Ireland will follow the example in the UK where there is an increasing focus on the use of data in journalism.

Following the interviews, some of the some of the better-known offerings in journalism were reviewed, such as the hybrid degree in computer science and journalism currently being offered by Columbia University, USA. This prompted a review of offerings internationally to see how European institutes are reacting to the emergence of computational journalism in the industry. An invaluable resource was found in the Italian Data journalism blog (http://datajcrew.sudmediatika.it) who maintain a listing of worldwide courses on their site. The data was taken into Ms Excel and collated by country to find that there are only 18 known courses worldwide.

The offerings are still few but this was expected given the topical nature of the subject. The USA and the UK lead in response to the needs to bring education into line with the new requirements for broader-skilled journalists but still with very few courses - 5 and 3 respectively. The US had embraced sophisticated coding techniques more in its courses.

In a survey of statistics education in UK Journalism in 2014 the Royal Statistical Society (Kemeny, 2014) found that not one of the 374 undergraduate courses reviewed offered statistics as a core module, though when asked course directors did say that they were reviewing their offerings.

The participant’s responses are in line with the literature on this topic. Utts has posited that the media are increasingly using statistics to communicate with the public and there is a need to ensure that practitioners are educated in their correct use. She suggests that the education approach

Findings and Analysis 37
should be to teach the rudiments of analytical thinking rather than focusing in on difficult theory (Utts, 2010). This last point is echoed by De Maeyer et al who noted that journalists are uncomfortable with maths and this feeds into the level of adoption of the tools and techniques of data journalism (De Maeyer et al., 2014).

**Finding** - the content to be added to existing courses is not highly technical in nature and does not stray into the computer science realm for the most part.

Responses from the participants on the nature of the technological skills needed indicate that educators are not required to incorporate difficult mathematical and coding modules. Generally, the suggested skills were more of the nature of better general online analysis skills, google search skills, video and audio analysis, familiarity with mapping tools and, above all, MS Excel.

One of the participants had also taken the online data driven journalism course offered by Canvas Network that was taken by the researchers in preparation to the study and she felt that that the course content would make an excellent starting point for integrating content into existing journalism courses.

### 4.7.2 Roles

A number of literature sources were found that analysed the changing role of the journalist due to increased technology in the industry. The researchers were particularly interested in asking key participants, particularly those currently working as journalists, if they were aware of a need to adapt their existing roles to embrace a more technical angle.

The following question was asked: Do you believe that increased use of technology in journalism is changing the traditional role of the journalist?

There was consensus among the participants that the role of the journalist has become a more hybrid role that encompasses technical skills. One participant felt that where once the roles of the journalist and the programmer were far apart, there was much more overlap now. She commented that the journalist now plays an important role as a data scientist with distinct responsibility for accuracy of data. Despite the evolving role description, the core traditional competencies were also vital – there was a sense that some of the older generation are scared of the idea of data journalism but in actual fact the two elements combined should make journalism better.

Another participant pushed for a much more defined redefinition of the role of the journalist in the data age. He compared journalists with the business analysts in a company that trawls for business information in a model not unlike Storyful. Claiming that the analysts do much the same work as journalists, he wondered if it was any longer appropriate to accentuate the difference:

> “*What’s the difference? And why do we have this distinction between information analysts? If the output is the only difference, the news discovery and information management is actually the same problem*”.

The next question was asked to tease out whether existing resources could cross-train into the area or whether the newsroom needed to accommodate journalists working alongside technical resources: As technologist make their way in journalism, do you think this will result in new roles with both skills or the separation of roles is more beneficial?

The consensus was that a mix of roles would be more beneficial, given the range of diverse skills that can be found in a newsroom:
“there are some people who are excellent with sources, there are some people who are brilliant with FOI, there are some people who are fantastic feature writers or brilliant news-people who couldn’t write a colour piece in a million years so I don’t think everyone is going to have every requisite skill.”

One participant brought it down to the lack of a common language. She felt there was a “divide between journalists and technology people. Technology people don’t talk the language of journalists and journalists don’t like to hear the language of technology people”.

Communication between the two groups needs to be simplified, particularly the interaction from technologists. On the other hand, journalists are slow to understand the nature of the technology which can support them in their job, echoing Emily Bell’s comment about the difficulty finding that rare creature that is both a journalist and a programmer (Lafrance, 2013).

In order to identify if the industry is impacted by the lag in integrating computational skills in current journalism courses, the following question was asked: Are you having any issues recruiting staff with the right mix of journalism & technical skills?

Generally, this came down to whether the right questions were being asked of candidates in order to select the right match. Are they technologists or journalists or in-betweeners?

Finding: There is little evidence of fear of change

Participants were asked if they were worried or excited about the future of journalists and the response was very positive. In general, respondents felt that the increased growth in data journalism made them feel like their futures were more secure. This is not to ignore the on-going issues in the industry as it struggles to marry print and online media. Respondents felt the growth of computational journalism was part of the solution to those issues.

The researchers would like to caution that a more representative and larger sample should be taken to truly test this finding, particularly as the candidate selection process targeted journalists already working in data driven journalism.

Finding: potential changes in employment terms may impact on innovation in the industry

In addition to changes in roles due to the increased use of technology, the changing nature of the employment model was also considered. This was prompted by many of the views of one of the key participants who viewed computational journalism more as a means of achieving efficiencies, and data journalism as a means of monetising a product. This participant is self-employed and would by default have different concerns to staff journalists. He commented that journalists by their nature are not very entrepreneurial and he bucks that trend being involved in start-ups and innovative product development.

From the review of the CSO Census 2011 figures, it was noted that there were changes in the employment profiles of journalists since the previous census in 2006 (Appendix I). While there was a 2% drop in staff journalists there was a 15% rise in self-employed journalists. The overall number of journalists employed in 2011 rose 10%. Whether these changes are as a result of the recent recession or due to the increased use of technology in journalism is outside the scope of this study.

However, the authors would postulate that a large swing towards self-employment in journalism may inhibit the future growth of computational journalism as freelancers may not have the time or resources to cross train on new techniques and tools and would not have the time to dedicate to uncovering data driven stories. This was indicated by J1 who does much of his most innovative work out of hours. He commented that other journalists have indicated to him that pressures of time mean they do not have the capacity to investigate new processes and concepts.
On the other hand, the proliferation of new tools and open source products has potentially lifted the barriers of entry to the industry.

**Finding: Role changes may be possible but need a community and industry support to bring this into the mainstream**

The participants interviewed for this study have developed a support system and community through which they share knowledge. To date the growth of data driven journalism in Ireland is as a result of their enthusiasm and hard work. There is little structural and industry support in place to aid them, though that is beginning to change. They can only achieve so much in the absence of a more widespread industry decision to embrace data journalism.

### 4.8 Regulatory Framework

As part of the process to chase up on some themes that emerged during the interviews with the key participants, the researchers interviewed Peter Feeney, the press ombudsman of Ireland.

Much of the findings documented in this section are outside the scope of the discussion in this report but they have been documented them in the conclusions chapter as worthy of further research.

As a background, the role of the press ombudsman is to deal with complaints from the public against the press. Membership is voluntary but all national newspapers are members and must abide by a Code of Practice. The office of the ombudsman does regulate the online version of a newspaper, including videos, but the user comments on an article are outside the remit of the Press Council. There is also a divide between print and broadcast elements, the latter being regulated by the Broadcast Authority of Ireland (BAI).

The following issues were highlighted during the key participants’ interviews and some detail was provided by the ombudsman.

**Outdated Code of Practice**

The code as it exists currently reflects a more static period and needs to be updated to take into account the increase in digital and online journalism. The technology is moving faster than the philosophy - as the platforms multiply and as the distinction between broadcasting and non-broadcast, between print and online disappears, an argument grows for a single code or regulation to cover all the disparate parts of new journalism. There is an added complication which is that issues tend to be multinational as data spans borders online and each country is scrambling independently to regulate on that. Some countries have implemented a single regulator. As the platforms multiply and as the distinction between broadcasting and non-broadcast, between print and online disappears, you can see the argument for one Regulator.

The ombudsman took the approach that journalism is journalism, whether it’s a newspaper, broadcast or online and therefore the standard should apply to all of it. But the limits of the office
plus the delineation of the different facets among multiple regulators make this a complex proposition in reality.

### Regulatory Gap

The Press council does not cover many of the areas that are an increasing source of complaints. There is an issue with the degree of information now kept online and a question to be asked here in the absence of a regulatory body with the right to protect the consumer - when does this level of information stop empowering the individual and start limit him/her? This issue arose consistently in the participant interviews on the subject of the digital archive.

There is a growth in complaints about on-line material, and the office gets complaints about companies like Facebook which, being a closed medium of communication (in that it doesn’t broadcast or publish) isn’t subject to the control of the regulator. Many of the complaints relate to the unauthorised use of photos from a profile but the Press council can currently do nothing about it. Twitter is also not publishing in the sense that it is generally understood and is also exempt.

The main area of concern for the Press Council is the fact that online comments are unregulated. The ombudsman would like to see comments treated similarly to letters published on the newspaper site and subject to the rule that, though the comments aren’t pre-moderated if a complaint is received and the comment is not reviewed or removed, it should come under the ombudsman.

The ombudsman made particular reference to a matter that the researchers had encountered during the literature review, which was that increased use of technology is diminishing or removing ethical barriers between the advertorial and the editorial (Pavlik, 2000). The ombudsman highlighted that it is sometimes difficult to tell if an item is news or an advertisement.

### 4.9 Potential Legal Issues

#### 4.9.1 Copyright

The participants were all asked: Do you foresee any copyright issues with the increased use of proprietary content and algorithms in online journalism?

Computational journalism, and data driven journalism specifically implies the use of information that might have been generated by users. YouTube is a large part of the content monitoring that Storyful do and their staff are well versed in copyright requirements.

The responses were that generally they all thought that copyright and the control and protection of copyright was a major issue. Some news outlets however don’t seem to accept the user as a partner when it comes to user generated content. They try to use the material without acknowledging the legitimate copyright owner and embed the video on their own site and use it to generate traffic. This is news “cannibalisation” (taking material from other news organisations, without attribution).
T1 talked about a constant battle to maintain copyright over content – when information is being found and processed that quickly, and is being disseminated by multiple sites over multiple servers and applications, it is difficult for the law protect the source of copyright.

Storyful makes the copyright owner a partner in the process – their value proposition is based on the 3 Cs (Credit, compensation and copyright) and the copyright owner is respected and compensated.

In analysing the general issue of copyright with Patricia Sheehy Skeffington, the researchers asked this question: Is the current law on copyright sufficient to deal with the new challenges that may be posed by computational journalism, in particular the use of algorithms to generate news articles?

The response was that it depends on the amount of control and input that derives from one and the other but authorship is something that is a matter that can be assigned by contract so the use of article generating software could be dealt with by contract. There is a grey area so it would be best to have a contract.

4.9.2 Data Protection

The questions that were asked by the researchers in this respect centred around the lack of transparency around the use of technology and algorithms in their processing – while the inputs may be known, the processing of those inputs are in a black box and there were questions around whether this could raise legal, as well as ethical, issues.

The question asked was: does ‘the public good’ exemption in the Data Protection laws apply to machine code as it does to a living journalist?

There is an exemption within data protection for journalism. While the public do have a right to know but that is subject to certain limitations and they are the privacy limitations, limitations in respect of defamation, limitations in respect of privacy.

In respect of the use of algorithms and the lack of transparency around the end to end of the process, it was suggested that the important thing was to set the contours of the algorithm ensure that it is kept within the boundaries of the law in all its processing.

The question that was asked was: Could incorrect code or bad parameters fed into an algorithm lead to a defamation case due to false information being published by a person or entity?

Patricia said it was absolutely possible and indeed had already happened. It was tied somewhat to the right to be forgotten as digital information remains accessible for a long time. The ombudsman also agreed with this and it this issue had been a source of consumer complaint.

“Access to information has changed everything. It has empowered people as it has given them much more information and for consumers it’s great … but I’m a 12 year old boy starting in a new school and my friends Google my name and up comes my dad’s conviction for crime 10 years earlier, does that empower me?”

4.10 Conclusions

The interviews provided a quality overview of the current state of computational journalism in Ireland. There were some interesting findings on the impact of technology and the analysis also uncovered some additional sub-themes that have been highlighted in the conclusions chapter of this document as areas of future research.
Chapter 5 Conclusion

The research described in this report recounts a rewarding journey for the researchers who had set out to discover if Irish journalism is becoming increasingly automated, and if this automation is in fact occurring, then would it imply a limit?

Given the broad approach taken by the study, the findings were comprehensive across a number of topics of interest. The researchers were interested to note that some of the participants were focused on the potential to create solution products to long standing challenges in the journalism pipeline, namely the sourcing and gathering of data to create news. This was closely allied to the conclusion that the article is no longer the only output to the journalistic workflow. The integration of data in the workflow has been a significant agent for change in the industry. The acknowledgment of the need to manage it, and the capacity to monetise data, has been one of the most obvious signals to the researchers that some principles of technological theory are being integrated into journalism.

The use of the term ‘computational journalism’ is not widespread in Ireland and the research highlighted the fact that it has become synonymous with data driven journalism. There has been a low level of data journalism in the country for the past few years and it has been practiced by a small number of advocates targeting data sources from public bodies and government. In the past year however, this has spread to some other organisations such as Insight Analytics and the Irish Times data blog so it can be expected that more outlets will follow.

The researchers were intrigued to uncover that data driven journalism actually invites the public into the role of the watchdog. In general, as the researchers have found from analysis of websites and from a review of the literature, data journalists are open with the data and methodologies they use to build their stories and actively invite readers to comment or use the data for further analysis. This collaborative ‘open source’ approach invites the public into the democratic process that has contradicted some of the assumptions in the current literature.

In order to support this growing movement towards the adoption of data driven journalism, educators need to incorporate the skills necessary to work with data. The researchers found that, in common with many other countries, Irish education courses are not preparing new journalists with the requisite skills. The study found that the level of technical skills needed is actually very low so it would not be difficult to include them. There is no need to include complex computer science theory for the main part – the research found that in fact a degree of delineation between the roles of coders and journalists is actually beneficial.

The study found a sense of optimism around the new employment possibilities around the increased use of data journalism that was not expected. The participants saw the increased incursion of technology into journalism as a solution, not a contributor, to the current debate around the need to re-evaluate the traditional business model in journalism.

However one finding that the researchers focused on was the possibility that a move towards self-employment in the industry might act as inhibitor to further innovation. Freelancers tend not to have the time or resources to dedicate to finding new ways of working and sourcing the news; rather the focus tends to be about creating output so they can get paid. The majority of the truly innovative work being done on data in Ireland has been by journalists working in their spare time almost as a labour of love. This isn’t sustainable if technological innovation is to be encouraged. A more comprehensive evaluation of this finding was beyond the scope of this study but the researchers believe it could be an interesting topic for future study.
Ultimately the innovation and automation in Irish journalism to date has centred on a small community of people who have looked for technological solutions to problems. To capitalise on this, a more systemic approach is required, involving all the players in the industry. Investment is needed.

Returning at last to the question posed by this study:

“Stop the Press! Or Press the Stop? A preliminary study of the convergence of technology and journalism in the Irish Context and its limitation”

The researchers can conclude that the presses will continue running for a while longer.

The researchers found there is a broad welcome for a significant amount of automation in Irish journalism. There is a reluctance to consider the automation of the creative process, namely the creation of an article, and it emerged that the study participants felt that it would take many years for the quality of the software to advance to the point where this would be a significant disruptor for the industry.

Finally, as supported by some of the literature, there is a belief that the intrinsic value and talent of a journalist cannot be replaced by artificial intelligence. Rather, as an industry, improvement can only come by man and machine working together.

5.1 Areas of future research

Conducted as a preliminary study of Computational Journalism in Ireland, the research has naturally developed on a broad scope and a number of perspectives had to be necessarily considered. As a result, due to the breadth of the research context a number of areas of future research have been identified. The recommendation is that these areas of future research can help develop computational journalism in Ireland as to equip the journalist with the tools that can empower him/her to fulfil his or her mission to contribute to democracy.

5.1.1 A more comprehensive study of the journalism industry

A suggestion that a more comprehensive and representative survey is undertaken to assess attitudes to increased automation through technology.

5.1.2 Digital Archiving

There is an argument to create a central data repository so that news can be annotated and contextualised rather than the current process of going through the process of accessing data, creating a story and then discarding the source.

5.1.3 Regulatory and legal gaps

The current code of practice and the structure whereby three different regulatory bodies – the Press Council, the Broadcasting Authority of Ireland and the Advertising Authority – regulate online content is possibly no longer sufficient. A new framework may have to be considered.

5.1.4 Copyright law

A study into the potential copyright and defamation issues that may arise from increased algorithms in journalism under European and Irish law would be a welcome addition to the literature.
5.1.5 Employment

Is technology responsible for the increase in the number of self-employed journalists since 2006? Or is it the recession? Will this limit innovation?

5.1.6 Automated writing

Is the Irish public ready for automated writing? A study whereby members of the public are presented with two articles – one written using software and one written by a human – and asked to evaluate the writing styles of both to identify if the public can identify automated writing.
Appendices

Appendix I – CSO Census 2011 Table 2

<table>
<thead>
<tr>
<th>Occupational group</th>
<th>At Work</th>
<th>Unemployed (incl. looking for first regular job)</th>
<th>Unemployment rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>212 Electrical and electronic engineers</td>
<td>5</td>
<td>884</td>
<td>12.7</td>
</tr>
<tr>
<td>215 Chemical, production, planning and quality control engineers</td>
<td>3</td>
<td>325</td>
<td>5.8</td>
</tr>
<tr>
<td>216 Design and development engineers</td>
<td>2</td>
<td>142</td>
<td>8.4</td>
</tr>
<tr>
<td>219 Other engineers and technologies n.e.c.</td>
<td>2</td>
<td>407</td>
<td>8.9</td>
</tr>
<tr>
<td>230 Laboratory technicians</td>
<td>3</td>
<td>440</td>
<td>8.1</td>
</tr>
<tr>
<td>231 Engineering technicians</td>
<td>2</td>
<td>295</td>
<td>10.1</td>
</tr>
<tr>
<td>232 Electrical and electronic technicians</td>
<td>1</td>
<td>568</td>
<td>8.8</td>
</tr>
<tr>
<td>233 Architectural, town planning, building and civil engineering technicians</td>
<td>1</td>
<td>655</td>
<td>26.3</td>
</tr>
<tr>
<td>239 Other scientific technicians n.e.c.</td>
<td>2</td>
<td>530</td>
<td>9.9</td>
</tr>
<tr>
<td>310 Draftspersons</td>
<td>2</td>
<td>527</td>
<td>23.7</td>
</tr>
<tr>
<td>312 Building inspectors and quantity surveyors</td>
<td>1</td>
<td>552</td>
<td>18.1</td>
</tr>
<tr>
<td>313 Marine, insurance and other surveyors</td>
<td>-</td>
<td>1</td>
<td>1.4</td>
</tr>
<tr>
<td>212 Health and related workers</td>
<td>34</td>
<td>2,998</td>
<td>2.8</td>
</tr>
<tr>
<td>220 Medical practitioners</td>
<td>5</td>
<td>183</td>
<td>1.5</td>
</tr>
<tr>
<td>221 Pharmacists, pharmacologists, ophthalmic and dispensing opticians</td>
<td>3</td>
<td>70</td>
<td>1.5</td>
</tr>
<tr>
<td>223 Dental practitioners</td>
<td>2</td>
<td>41</td>
<td>2.1</td>
</tr>
<tr>
<td>224 Veterinary</td>
<td>5</td>
<td>30</td>
<td>1.7</td>
</tr>
<tr>
<td>230 Nurses and midwives</td>
<td>64</td>
<td>1,115</td>
<td>2.9</td>
</tr>
<tr>
<td>231 Medical radiographers</td>
<td>7</td>
<td>21</td>
<td>1.3</td>
</tr>
<tr>
<td>231 Physiotherapists and diagnosticians</td>
<td>2</td>
<td>70</td>
<td>2.1</td>
</tr>
<tr>
<td>233 Physiotherapists and diagnosticians</td>
<td>2</td>
<td>70</td>
<td>2.1</td>
</tr>
<tr>
<td>234 Other health associate professionals n.e.c.</td>
<td>2</td>
<td>88</td>
<td>2.0</td>
</tr>
<tr>
<td>235 Other health associate professionals n.e.c.</td>
<td>17</td>
<td>361</td>
<td>4.8</td>
</tr>
<tr>
<td>236 Nurses aids</td>
<td>11</td>
<td>67</td>
<td>6.0</td>
</tr>
<tr>
<td>237 Social workers and related occupations</td>
<td>36</td>
<td>1,380</td>
<td>6.1</td>
</tr>
<tr>
<td>238 Psychologists and other social-behavioural scientists</td>
<td>9</td>
<td>316</td>
<td>11.7</td>
</tr>
<tr>
<td>239 Social workers and probation officers</td>
<td>9</td>
<td>152</td>
<td>3.8</td>
</tr>
<tr>
<td>240 Social workers and probation officers</td>
<td>20</td>
<td>912</td>
<td>5.7</td>
</tr>
<tr>
<td>241 Social workers and related occupations</td>
<td>43</td>
<td>83</td>
<td>1.9</td>
</tr>
<tr>
<td>242 Religious occupations</td>
<td>43</td>
<td>83</td>
<td>2.3</td>
</tr>
<tr>
<td>4726 Other professional workers</td>
<td>132</td>
<td>6,029</td>
<td>12.8</td>
</tr>
</tbody>
</table>

**Note:** The table provides data on various occupational groups and their employment status as recorded in the CSO Census 2011.
## Appendix II – CSO Census 2011 figures

<table>
<thead>
<tr>
<th>Year</th>
<th>Total</th>
<th>Employer or own job</th>
<th>Employee Account</th>
<th>Employee assisting Relative</th>
<th>Total Employed</th>
<th>Unemployed (incl. looking for first regular job)</th>
<th>Unemployment Rate</th>
<th>Occupational group</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006</td>
<td>6096</td>
<td>1910</td>
<td>3962</td>
<td>11</td>
<td>5883</td>
<td>96.51%</td>
<td>213</td>
<td>3.5</td>
</tr>
<tr>
<td>2011</td>
<td>6721</td>
<td>2188</td>
<td>3887</td>
<td>18</td>
<td>6093</td>
<td>90.66%</td>
<td>628</td>
<td>9.3</td>
</tr>
</tbody>
</table>

% shift: 10% 15% -2%

4% -6% 195% 166%

---

**Note 1:** Census 2011 Table 1 shows a 'Total in Labour' figure of 2,232,203.

**Note 2:** Notice that on page 13 of the report, the figure in the occupation is 6093 but in the subsequent tables it's 6721.

**Note 3:** Table 1 (overall employment stats by industry) there is no category for occ code 380.
Appendix III – Interview Questions

<table>
<thead>
<tr>
<th>Interview Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do you think that the way the news is told will be different with the increased use of algorithms and data analytics?</td>
</tr>
<tr>
<td>In your opinion, do you see automated journalism as an enemy or an ally of the modern journalist?</td>
</tr>
<tr>
<td>Would you, or do you currently, use software to write or source your stories?</td>
</tr>
<tr>
<td>Are you excited or concerned about the journalism industry in Ireland in the future?</td>
</tr>
<tr>
<td>Are you having any issues recruiting staff with the right mix of journalism &amp; technical skills?</td>
</tr>
<tr>
<td>As technologist make their way in journalism, do you think this will result in new roles with both skills or the separation of roles is more beneficial?</td>
</tr>
<tr>
<td>Do you believe that increased use of technology in journalism is changing the traditional role of the journalist?</td>
</tr>
<tr>
<td>Do you foresee any copyright issues with the increased use of proprietary content and algorithms in online journalism?</td>
</tr>
<tr>
<td>Do you see the increasing conjunction of technology and journalism in Ireland?</td>
</tr>
<tr>
<td>Do you think that digital journalism can develop in tandem with traditional print media models?</td>
</tr>
<tr>
<td>Do you think that journalists and journalists’ educators are equipped with the right tool set and skills to capitalise on the increased use of technology?</td>
</tr>
<tr>
<td>Do you think that the way the news is told will be different with the increased use of algorithms and data analytics?</td>
</tr>
<tr>
<td>Do you think the increased use of analytic investigation and the graphic visualisations of results are good methods of engaging with readers?</td>
</tr>
<tr>
<td>How are you currently using technology in your job? What percentage? What tools?</td>
</tr>
<tr>
<td>How are you currently using technology in your writing?</td>
</tr>
<tr>
<td>How far do you see automation going?</td>
</tr>
<tr>
<td>How is information gathering going to change now that there is access to enormous amounts of data - more challenging? More accurate? Less accurate?</td>
</tr>
<tr>
<td>How would you define Storyful, is it a technology firm or a news firm?</td>
</tr>
<tr>
<td>If data journalism can be considered an automation of the journalistic workflow (gathering) - can you see automating writing as the next step?</td>
</tr>
<tr>
<td>In your opinion what is the current level of adoption of automated journalism in Ireland?</td>
</tr>
<tr>
<td>What advanced technology do you use in each of the step of the journalistic workflow - Gathering, Producing, Distributing,</td>
</tr>
<tr>
<td>What is the ratio technical vs journalists?</td>
</tr>
<tr>
<td>You use news generating algorithms and proprietary software to - do you use any news generating or content writing algorithms?</td>
</tr>
</tbody>
</table>
## Appendix VI – List of Data Driven Journalism Courses offered worldwide

Note: sourced on http://datajcrew.sudmediatika.it

<table>
<thead>
<tr>
<th>Country</th>
<th>Institution</th>
<th>Course Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>Melbourne University</td>
<td>data Journalism course</td>
</tr>
<tr>
<td>BsAs, Argentina</td>
<td>ETER School of Communication</td>
<td>Postgraduate course on New contents for digital platforms.</td>
</tr>
<tr>
<td>Canada</td>
<td>University of King’s College, Halifax</td>
<td>Summer School in Data Journalism</td>
</tr>
<tr>
<td>China</td>
<td>University of Hong Kong</td>
<td>Data Journalism Scholarships, Journalism and Media Studies Centre</td>
</tr>
<tr>
<td>France</td>
<td>École de Journalisme de Sciences-Po</td>
<td>Training course in “numerical journalism” - French idiom of Data Journalism</td>
</tr>
<tr>
<td>Italy</td>
<td>The Data Journalism Working Group at MediaLab Prado</td>
<td></td>
</tr>
<tr>
<td>Netherlands</td>
<td>Tilburg University</td>
<td>Master in data Journalism</td>
</tr>
<tr>
<td></td>
<td>University of Amsterdam</td>
<td>data visualization course, MA in Media studies</td>
</tr>
<tr>
<td>South Africa</td>
<td>The University of the Witwatersrand in Johannesburg</td>
<td>Data Journalism Course, Wits University.</td>
</tr>
<tr>
<td>Spain</td>
<td>The School of journalism and communication of Madrid</td>
<td>Master in Investigative and Data Journalism, and Visualization</td>
</tr>
<tr>
<td>UK</td>
<td>Birmingham City University</td>
<td></td>
</tr>
<tr>
<td>Master in online Journalism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The Centre for Investigative Journalism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summer School</td>
<td></td>
<td></td>
</tr>
<tr>
<td>city university of London</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Master in Interactive Journalism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>USA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Columbia Journalism School</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dual degree: Journalism and Computer Science</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stanford University.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Graduate Program in Journalism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arthur L. Carter Journalism Institute, New York University.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>part of The Bachelor in Journalism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missouri School of Journalism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computer-Assisted Reporting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>School of Information, University of Michigan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>an Undergraduate degree and a Master of Science in Information sector</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2 Data Driven Courses
References


References

52


UTTS, J. 2010. UNINTENTIONAL LIES IN THE MEDIA: