Is Operational Research in UK Universities 'Fit-for-Purpose' for the Growing Field of Analytics?

By Michael J. Mortenson

Loughborough University, School of Business & Economics

DOCTORAL THESIS

This dissertation is in partial fulfilment of the requirements for the award of Doctor of Philosophy of Loughborough University

Michael J. Mortenson – June 2018

This work is dedicated to the memory of my MSc dissertation supervisor, Mr Vernon Blackmore, and PhD supervisor Professor Neil Doherty. Their guidance, contributions, and expertise have had significant influence on this work, and on my skills as a researcher, and on my life as whole.

I am eternally grateful to both of them, and hope this work can be, in some small way, some tribute to their kindnesses, hard work and influence.

DECLARATION

This dissertation is the result of my own work and includes nothing, which is the outcome of work done in collaboration except where specifically indicated in the text. It has not been previously submitted, in part or whole, to any university of institution for any degree, diploma, or other qualification.

Signed:

Date: 19th September 2017

By Michael J. Mortenson (BA (Hons), MSc, CAP)

Loughborough University

ABSTRACT

Over the last decade considerable interest has been generated into the use of analytical methods in organisations. Along with this, many have reported a significant gap between organisational demand for analytical-trained staff, and the number of potential recruits qualified for such roles. This interest is of high relevance to the operational research discipline, both in terms of raising the profile of the field, as well as in the teaching and training of graduates to fill these roles. However, what is less clear, is the extent to which operational research teaching in universities, or indeed teaching on the various courses labelled as "analytics", are offering a curriculum that can prepare graduates for these roles.

It is within this space that this research is positioned, specifically seeking to analyse the suitability of current provisions, limited to master's education in UK universities, and to make recommendations on how curricula may be developed. To do so, a mixed methods research design, in the pragmatic tradition, is presented. This includes a variety of research instruments. Firstly, a computational literature review is presented on analytics, assessing (amongst other things) the amount of research into analytics from a range of disciplines. Secondly, a historical analysis is performed of the literature regarding elements that can be seen as the pre-cursor of analytics, such as management information systems, decision support systems and business intelligence. Thirdly, an analysis of job adverts is included, utilising an online topic model and correlations analyses. Fourthly, online materials from UK universities concerning relevant degrees are analysed using a bagged support vector classifier and a bespoke module analysis algorithm. Finally, interviews with both potential employers of graduates, and also academics involved in analytics courses, are presented.

The results of these separate analyses are synthesised and contrasted. The outcome of this is an assessment of the current state of the market, some reflections on the role operational research make have, and a framework for the development of analytics curricula.

The principal contribution of this work is practical; providing tangible recommendations on curricula design and development, as well as to the operational research community in general in respect to how it may react to the growth of analytics. Additional contributions are made in respect to methodology, with a novel, mixed-method approach employed, and to theory, with insights as to the nature of how trends develop in both the jobs market and in academia. It is hoped that the insights here, may be of value to course designers seeking to react to similar trends in a wide range of disciplines and fields.

Keywords: Operational research; analytics; big data; data science; curricula development

ACKNOWLEDGEMENTS

My sincere thanks and gratitude to Professor Stewart Robinson and Professor Neil Doherty for supervising this research and guiding me through this process. They frequently went above and beyond the call of duty for an academic supervisor, and their efforts not only have had a significant and positive impact on this work, but also my career and life in general. Elsewhere, my thanks too to everyone I interacted with in the School of Business and Economics, and across Loughborough University as a whole. The university's reputation for student experience is well deserved, and it was a pleasure to have worked and interacted with everyone there.

Finally, I would also like to thank the sponsors of this research, the O.R. Society, and my (long suffering) wife and family for their support and patience through this process that end up being far longer than I could envision.

Thank you to Kayla Friedman and Malcolm Morgan of the Centre for Sustainable Development, University of Cambridge, UK for producing the Microsoft Word thesis template used to produce this document.

CONTENTS

Michael J. Mortenson – June 2018

List of Tables

LIST OF FIGURES

LIST OF ABBREVIATIONS AND ACRONYMS

ACM – Association of Computing Machinery

AIS – Association of Information Systems

ANOVA – Analysis of Variance

API – Application Program Interface

B2B – Business to Business

B2C – Business to Customer

BA – Business Analytics

BI – Business Intelligence

CAP – Certified Analytics Professional

CLR – Computational Literature Review

CRM – Customer Relationship Management

CS – Computer Science

CSS – Cascading Style Sheets

DFS – Distributed File System

DOE – Design of Experiments

DSS – Decision Support Systems

DTM – Document-Term Matrix

ERP – Enterprise Resource Planning

EM – Expectation Maximization

FTP – File Transfer Protocol

GUI – Graphical User Interfaces

HDFS – Hadoop Distributed File System

HR – Human Resources

HTML – Hypertext Mark-up Language

IEEE-CS – Institute of Electrical and Electronics Engineers – Computer Society

IFORS – International Federations of Operational Research Societies

INFORMS - The Institute for Operations Research and the Management Sciences

KDD – Knowledge Discovery in Databases

KPI – Key Performance Indicator

IoT – Internet of Things

IPA – Interpretative Phenomenological Analysis

IS – Information Systems

IT – Information Technology

LDA – Latent Dirichlet Allocation

LP – Linear Programming

LSA – Latent Semantic Analysis

LSTM – Long Short-term Memory

MATLAB – Matrix Laboratory

MCDA – Multi-Criteria Decision Analysis

MIS – Management Information Systems

ML – Machine Learning

MMR – Mixed Methods Research

MPP – Massively Parallel Processing

MSc – Masters of Science

MTW – Module Topic Weighting

OLAP – Online Analytical Processing

OR – Operational Research

OR/MS – Operational Research/Management Science

PCA – Principal Component Analysis

PL/SQL – Procedural Language/Structured Query Language

pLSI – Probabilistic Latent Semantic Indexing

PSM – Problem Structuring Method

RFID – Radio Frequency Identification

RNN – Recurrent Neural Network

SQL - Structured Query Language

SVC – Support Vector Classifier

SVM – Support Vector Machine

TF-IDF – Term Frequency – Inverse Document Frequency

VBA – Visual Basic for Applications

VEM – Variational Expectation Maximization

LIST OF APPENDICES

Michael J. Mortenson – June 2018

1 ANALYSING ANALYTICS

In an evaluation of our question, “Is operational research in UK Universities ‘Fit-for-Purpose’ for the growing field of analytics?”, there are several distinct aspects. These include, but are not limited to, the relationship between analytics and operational research (OR), requirements of the analytics employment market, the current provisions in UK universities, and, as the most obvious starting point, the exact definition of this ‘growing field’. Therefore, this chapter will begin the study by evaluating prior research and commentaries on these topics, from which the research gaps and objectives can be drawn. As such, in contrast to a “standard” thesis, the structure slightly differs in that this chapter combines the traditional “introduction” and “literature review” sections into one. This is partly motivated by the breadth of the area under investigation, but also the benefit of this approach is that this allows us to present, in a more inductive fashion, the gaps and concerns that shape this thesis, as they emerge in the existing literature. Thusly a clearer justification of the approaches, topics and research objectives included can be presented.

To this end, the chapter is arranged as follows. We begin by highlighting key evidence of analytics’ growing importance in both practice and academia, and the relevance of this growth to the operational research discipline. This essentially provides the motivation and relevance for the work. Secondly, we compare and contrast definitions of analytics given in prior research, and some of the contradictions they present. Thirdly, we provide a general summary of previous work into analytics, highlighting the key themes discussed in the literature, and the most influential works, journals and authors. The fourth section explores analytics job roles, and the fifth provision of analytics education, comparing this to operational research in each case. The sixth section analyses literature related to curricula development, and some of the issues and potential barriers therein, before the final section summarises the research objectives and provides the structure of the remainder of the thesis.

1.1 Operational Research and the Growth of Analytics

As stated, we begin this chapter, and the thesis, by discussing the growth of analytics, fundamental to a justification for the research, as well as the context within which it is positioned. Fortunately, to this end, the evidence of analytics' growing importance is plentiful. One such example, albeit not necessarily the most scientific of approaches, and one which has obviously limitations in respect to the veracity of the data, is through examining Google search data. Figure 1 shows the relative frequency of searches for keywords related to the topic of analytics, to OR, and also business analytics from January 2004 to July 2017 (via Google Trends). As can be seen, there has been a significant growth over this time, with a significant spike at the end of 2005, and a consistently high search volume since 2009, an indicator of growing interest. By comparison, OR searches show negative correlation, declining when analytics searches increase. "Business analytics" topic searches are at a much lower rate so are hard to compare.

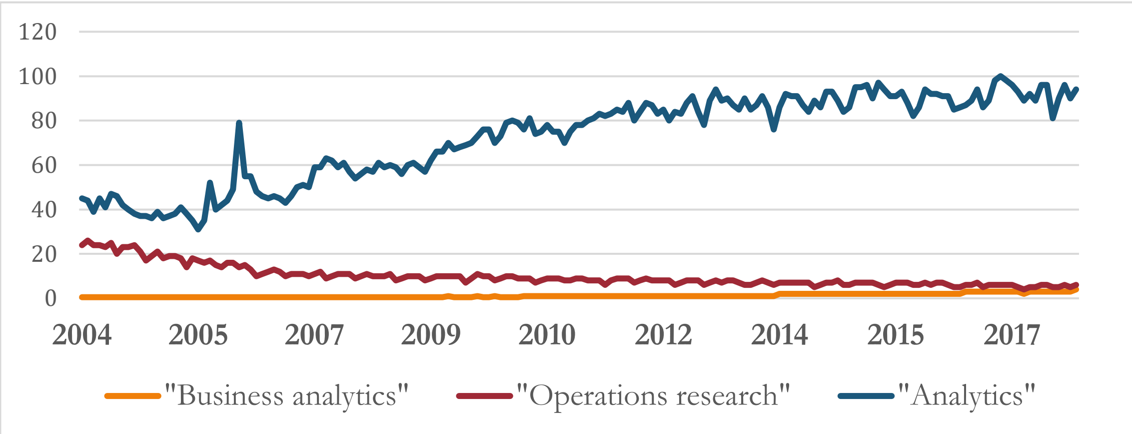


Figure 1 Google searches for topics related to analytics and OR 2004-2017

(Source: https://www.google.co.uk/trends)

Within the academic literature this growth is also apparent. Since 2012, four new journals with "analytics" in their titles have begun circulation (*Decision Analytics*, the *International Journal of Business Analytics*, the *International Journal of Business Analytics and Intelligence*, and the *Journal of Marketing Analytics*). Additionally, and within the same timeframe, a further seven journals with "big data" in their title, and five with "data science" (both terms which are often strongly associated with analytics) have been commissioned. Similarly, the number of academic publications concerning analytics has grown significantly within this time period. Chen *et al.* (2012) report the publication of 126 academic articles in business journals in 2011 containing the phrase "business analytics" in the title or abstract, equal to the total published in the ten years' prior (252 articles in total between 2000 and 2011).

The jobs market for analytics professionals, as will be discussed in more detail in section 1.4, is seemingly similarly buoyant. One of the most cited sources in this regard, is a McKinsey report forecasting a skills shortage (in the US alone) of between 140,000-190,000 "deep analytical talent positions" and 1.5million "data-savvy managers" by 2018 (Manyika *et al*, 2011). A similar report from e-Skills and software vendor SAS suggests a 243% increase in demand for "big data specialists" in the UK (e-Skills and SAS, 2013).

Finally, we can also track analytics' growth in respect to educational offerings. Researchers associated with a variety of academic disciplines, such as information systems (e.g. Chiang *et al*, 2012) and, of particular relevance to this study, OR (e.g. Liberatore and Luo, 2010; Ranyard *et al*, 2015), have pointed to the synergies between their discipline and analytics, and the potential of their degree courses to meet the aforementioned demand. Further, many universities have sought to meet this need through specialised degrees, with titles such as "Business Analytics" or "Data Science", which have proliferated throughout Europe, North America and the rest of the world. A recent report from Deloitte identifies over 100 analytics-related degree programs in the US alone (Danson *et al*, 2016).

The causes for this growth are likely to be complex and multi-faceted. Some of these are more explicit and easily identifiable, whereas others are more hidden and subtle. As inferred in some of the earlier discussion of this section, one potential stimulus is the increased availability of data in business and in society, mostly due to the ubiquity of the internet, and the ability to record, generate and share data at unprecedented levels. As illustration, Helbing and Balietti (2011) estimate that 1,200 exabytes of data were generated in 2011, compared to only 150 exabytes in 2005 (a 800% increase in only six years). The consultancy IDC estimated that in 2016 this had risen to 16.1 zettabytes (16,100 exabytes) and that by 2025 there will 163 zettabytes of data in the world (Reinsel *et al,* 2017).

A simplistic interpretation of this, with data representing an essential ingredient in most analyses, is that the greater the available data, the greater demand for professionals who can interpret it. Although simplistic, there is likely some truth to this. Hal Varian, Google's Chief Economist, pointed to this connection in an online question and answers session:

> "So what's getting ubiquitous and cheap? Data. And what is complementary to data? Analysis. So my recommendation is to take lots of courses about how to manipulate and analyze data: databases, machine learning, econometrics, statistics, visualization, and so on." (Freakonomics, 2008)

However, it is not just the volume of data being generated that may be relevant in this equation, but also the form and the subjects of the data itself. With so many people, devices and objects connected to the internet, data can be recorded autonomously and in places that were typically 'hard to reach' in traditional research.

For instance, in supply chains, the ability to attach RFID (radio frequency identification) chips to products means that companies can record in near real-time location information on products and inventory (e.g. Asif and Mandviwalla, 2005). In consumer research, previously data on customer opinions and preferences was collected, often at not insignificant cost, in the form of questionnaires or focus groups. With the proliferation of social media, review websites and user generated content, businesses can instead extract similar results from content published online. In both cases, however, there is requirement to work with 'non-traditional' data sources, which can present technical challenges. The analysis of RFID data requires filtering out 'noise' such as accidental movement of goods during shelf-stacking activities; the analysis of online customer comments requires often complex extraction and pre-processing steps (e.g. Liu and Zhang, 2012). In other words, it is not just that there is more data for analysis, but also that there is data available for analysis that was hard to obtain in a pre-internet world, with significant potential value to businesses, and which present new challenges in their management and interpretation.

Alongside such more overt changes, it is also worth considering the groups and organisations which may influence this growth in awareness and engagement with analytics. Such a trend does not spread without promoters or mediators; no-one searches Google for terms such as "analytics" and "business analytics" (figure 1), without some form of primer. In the case of analytics, the sources of such influence are likely varied, from those with 'something to sell' (such as consultancies, software vendors or even authors and researchers), through to the various media outlets.

In such circumstance, again there is likely some mixture of 'signal' and some of 'noise'. As detailed above, while there are tangible reasons why analytics should be growing in popularity, inevitably there is also hype. Mithas *et al.* (2013) describe the situation of big data as one of "a great deal of hype, confusion, and fear regarding big data, and numerous vendors have attempted to hijack the term for their own commercial benefit." Symbolic of the extent of such hype, even within this there are hype cycles, like matryoshka dolls, best exemplified by Gartner publishing an annual "Business Intelligence and Analytics Hype Cycle" since 2013 (e.g. Schlegel and Hare, 2017).

Growth alone, even where it can be distinguished from hype, is, however, only relevant to this research if it can be considered relevant to the OR discipline. To this regard, equally there is considerable evidence. Firstly, as explored further in the next section of the chapter, the definitions of analytics and OR generally used in the literature show significant communality. For instance, INFORMS' definitions of analytics as "the scientific process of transforming data into insight for making better decisions" (Liberatore and Luo, 2011); bears close relation to their definition of OR as "the application of advanced analytical methods to make better decisions" (INFORMS, 2013). One could argue that the former specifically mentions "data" while the latter does not, but "advanced analytical methods" clearly infers the use of data. An association is also made clear in some of the analytics degree programs offered in universities. For instance, the "Operational Research & Management Science" group of the University of Warwick is listed as delivering their MSc in "Business Analytics" (www.wbs.ac.uk/research/specialisms/teaching-groups/orms/).

Secondly, again a subject of discussion further into the thesis, there is some evidence that OR is currently facing significant challenges. In an analysis of the OR "ecosystem", Sodhi and Tang (2008) identify numerous weaknesses inherent to the discipline and several threats that it may face, including an unclear identity (particularly relevant in consideration of the similarity of the definitions given of the discipline and of analytics), issues with the journals and societies associated with the field, a declining position in university business schools and a lack of visibility in the jobs market. Whilst the study is now nearly a decade old, several other authors (e.g. Liberatore and Luo, 2010; Evans, 2012; Royston, 2013r eRanyard *et al*, 2015) point to the potential benefits analytics may bring to the discipline to these regards, thereby suggesting some of these problems may remain valid today.

As this discussion demonstrates, analytics is both seemingly a growing concern, and one which is suggested to have clear relevance to the OR discipline. However, it is often difficult to discuss the topic of analytics without also considering related terms such as "big data", "data science" and several others. Therefore, to evaluate the relationship between analytics and some of these terms, the next section will review some of the discussion in this area, as well as the definitions given by various authors.

1.2 Definitions of Analytics

In almost all studies of this kind, it is important to define one's terms and the meaning of key concepts. However, this is particularly relevant in this research due to ambiguities as to the precise definition of the term "analytics" and its overlap with related concepts. Perhaps the most cited definition of analytics is provided by Davenport and Harris (2007, p 7):

> "[T]he extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and actions. The

analytics may be input for human decisions or may drive fully automated decisions. Analytics are a sub-set of [...] business intelligence."

This definition, one which is more clearly elaborated than that of INFORMS (given in section 1.1), is sufficient to act as a reasonable working definition of analytics for the study, although this will be developed over the course of the work. However, that is not to say that it does not present some issues. The claim that analytics is a subset of business intelligence (BI) is a view supported by others such as Randy Bartlett (2013, p 4) who argues "Business Intelligence = Business Analytics + Information Technology". However, this is contradicted in other research: Vesset *et al.* (2012) and SAP (2013) state the opposite view, describing BI as the subset of analytics. The work of Chen *et al.* (2012), Chiang *et al.* (2012) and Lim *et al.* (2012) sidesteps this by considering the two as a composite, using the acronym "BI&A". The inference in their work is that the first part of the acronym refers to the technologies that process and manipulate data, and the latter its analysis. A more cynical perspective is that the distinction is essentially superfluous, and that discussion of "analytics" is effectively an attempt to reinvigorate interest in the existing field of BI (Eckerson, 2011; Elliot, 2010).

This ambiguity is not confined to the differences between analytics and BI; there are other examples where definitions of analytics can be seen to be very similar to other supposedly separate fields. For example, the definitions of both OR and analytics given in section 1.1 show significant similarity. Further, Laursen and Thourland (2010, p XII) define analytics as "delivering the right decision support to the right people at the right time". This definition is very similar to that given by Shim *et al.* (2002) to the field of decision support systems (DSS): "technology solutions that can be used to support complex decision making". Clear argument can be made that the definitions are somewhat interchangeable, ergo that partitions between each are ill-defined.

An alternative approach, popular in practitioner literature, is to define analytics not as a concept but as a practice. The most prevalent of such definitions is proposed in Lustig *et al.* (2010), who argue that analytics comprises of three distinct aspects:

1. Descriptive analytics: statistical methods designed to explore "what happened?"
2. Predictive analytics: machine learning methods designed to predict "what will happen next?"
3. Prescriptive analytics: (primarily) OR methods designed to answer "what should the business do next?"

Whilst this description has been widely cited (e.g. Johnson, 2012; Walker, 2012; Basa, 2013), there is no clear division between these practices and those that may be considered part of the related fields discussed. Descriptive (which can be read as a combination of information systems and basic statistics) and prescriptive analytics (OR) are clearly well-established disciplines (albeit renamed) that have long been used in business decision making. Predictive analytics, whilst regarded by many to be an evolution of the approaches of data mining and machine learning (Agosta, 2004; Shmueli and Koppius, 2011), still has such commonality with these disciplines as to make a complete distinction problematic.

Other discussions of analytics suggest alternate disciplines as providing the source material for analytics. Chiang *et al.* (2012) posit the key areas are "data management, database systems, data warehousing, data mining, natural language processing, [...] network analysis/social networking, optimization, and statistical analysis" and that practitioners are "able to understand business needs, interpret the analyses performed on big data and provide leadership for data-informed decision making". Varshney and Mojsilovic (2011), however, propose "applied mathematics, applied probability, applied statistics, computer science, and signal processing" whereas Evans (2012) argues for BI/information systems, statistics and OR.

Further evidence of this ambiguity is present elsewhere in the literature. An example of this comes from a survey of the membership of INFORMS, the US OR society (Liberatore and Luo, 2011). Asked what the relationship between OR and analytics was, whilst few thought they were the same thing or entirely distinct, there was little consensus as to whether one is the superset of the other, of if they are separate fields sharing an overlap, shown in figure 2.

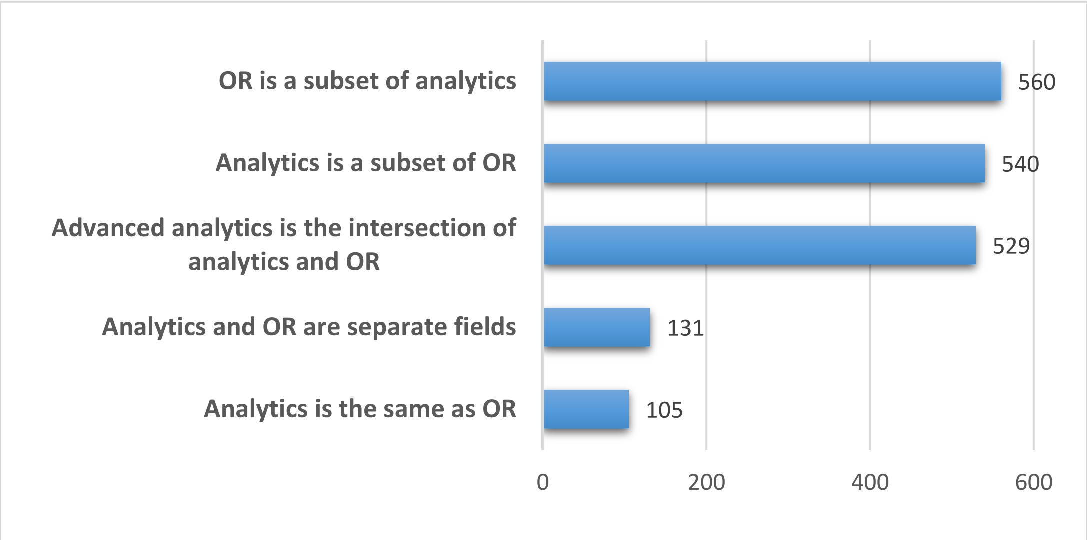


Figure 2 The relationship between OR and analytics

(Adapted from: Liberatore and Luo, 2011)

1.2.1 Analytics, Big Data and Data Science

Such ambiguity does not only exist in respect to analytics and pre-existing terms; there is also some lack of clarity as to how analytics overlaps with other newer terms, such as ‘data science’ and ‘big data’. This is evidenced in a recent paper into the area, whereby the authors group the terms “data science, predictive analytics, and big data” as a “confluence” they denote as “DPB” (Waller and Fawcett, 2013).

In respect to data science, there is already ambiguities inherent in the term. As Van der Aalst and Damiani (2015) note, amongst others, Danish computer scientist Peter Naur frequently used the terms as far back as the early 1970s. In this context, the term was used almost interchangeably with ‘computer science’. Discussion of the precise definition of data science is, again, limited and at time contradictory. Indeed, what may be the closest to a consensus view in the area, as given in Provost and Fawcett (2013, p 52), may be that “trying to define the boundaries of data science precisely is not of foremost importance. Data-science academic programs are being developed, and in an academic setting we can debate its boundaries”. Such a debate is even more complicated when we seek to draw boundaries with a similarly ‘new’ field such as analytics, a field that already has similar ambiguities (as discussed above) with existing fields and disciplines. There is an argument that the profession associated with this field, the ‘data scientist’, may be easier to separate from other analytics-type roles, a subject that will be discussed in section 1.5.

The term ‘big data’ is equally somewhat slippery and hard to define. This is perhaps in part because the name itself and its inference that is simply relates to large datasets. This is certainly not the consensus of the literature, or rather that this is not the whole story. The most famous description of the data that make up big data, although actually predating the popularisation of the term, is Gartner’s ‘three V’s’ model (Laney, 2001): albeit a model that has since been supplemented with many other ‘V’s by many other authors. The basic concept is that the datasets now being used differ from ‘traditional’ data in respect to their:

- Volume: containing millions of rows and in the order of gigabytes or terabytes;
- Variety: containing not only ‘structured’ data (such as financial records or closed-question survey responses) but also ‘unstructured’ data (such as text, audio or video);
- Velocity: data which is created or modified at great speed, accessible via APIs or data pipelines, and used in application where “the data is not the ‘stock’ in a data warehouse but a continuous flow” (Davenport *et al*, 2012).

This conceptualisation, though not strictly a definition, is probably the most used description of big data. However, it is not without issues. Firstly, volume, much like the name 'big' itself, is obviously relative. Long before the term was coined, many banks, retailers and governments used datasets large enough to be called 'big' my most people's description. Equally, with the amount of data being generated each day estimated in the quintillions of bytes (IBM, 2013), what is 'big' in today's terms is likely to be considered small in just a few years. Secondly, variety, and the term 'unstructured data', can potentially mislead. Essentially all data have structure, just we are more familiar with, and better equipped to deal with, some forms over others. A music file has clear structures in its tempo, pitch and decibels; images have colours, pixels and resolution; and the characters, white-space and word-counts of text files are as much a structure as present in 'traditional' data. Finally, velocity allows for many new applications and innovations for real-time analytics and offer opportunities in fields such as traffic management, disaster recovery and logistics. However, that is not to say every application of big data needs to be real-time; many analyses are performed weeks, months and even years after the data were collected. In summary, the perspective taken here of the three 'V's definition, is that presented by Mayer-Schönberger and Cukier (2013, p. 199), "useful for its time but imperfect".

The conceptualisation used in this study is moreover based upon big data's implications rather than the specific qualities of the data themselves, and is essentially two-fold. Firstly, this can mean data that is of such complexity as to make it difficult to manage in 'traditional' data ware-house-type architectures, despite all the benefits these can bring. This point is illustrated on the MIKE 2.0 (2012) website:

> "Big data doesn't lend itself well to being tamed by standard data management techniques simply because of its inconsistent and unpredictable combinations. A good definition of big data is to describe "big" in terms of the number of useful permutations of sources making useful querying difficult [...] and complex interrelationships making purging difficult."

The second difference between big and 'traditional' data is in the analytical methods entailed; there are, as Mayer-Schönberger and Cukier (2013, p. 6) describe, "things that one can do at a large scale that cannot be done at a smaller one". Large datasets, assuming that the data is of significant quality, make it possible to identify trends and patterns that may be too "weak" to be found in small samples, and approaches can be introduced which would not be possible without the scale of data (and processing power) now available.

A comparison between big data and analytics is in some ways simpler, as ultimately one relates to objects (data and data-types) and the other processes and methodologies. However, the almost simultaneous emergence of the terms (and that of data science), the frequency with which both topics co-feature in academic and practitioner literature, as well as the communality between the areas, is seemingly beyond mere coincidence. One convenient answer could be that "analytics" (or for that matter "data science") would simply translate as business intelligence, decision support systems and/or disciplines such as OR, but performed on big data. Whilst that would indeed make the task of this section much easier, this is not supported in the literature in that there are countless examples given of the use of "analytics" but on data that doesn't qualify as "big" on any of the criteria discussed above. To give one illustration of this, the influential *Competing on Analytics* book lists an early example of how analytics is used in the form of FICO scores being applied to predict the likelihood of "automobile accidents" (Davenport and Harris, 2007, p 26). Consequently, the relationship between analytics and big data remains fuzzy, as indeed are its relationships with many of the other fields discussed in this section.

1.2.2 Summary

Based on this discussion it remains unclear as to the precise definition of analytics, and how it relates to related terms, practices and disciplines; within which we would include the OR discipline. Rather than this being mere pedantry, for the nature of our task to have meaning, comparing the provisions of OR in universities to the needs of analytics, clarity as to what each of these elements means, becomes mission critical. Not only this, from a perspective of the literature and overall gaps, better illustrating the differences and associations between such terms is a useful and potentially important contribution. Therefore, we posit the first objective of the research as the following:

> **RO1:** To determine the relationship between academic definitions of analytics, operational research, and other related fields and disciplines.

Prior to addressing this, however, the remaining objectives need to be defined. In order to do so, the next section of this chapter will evaluate the analytics research area, and identify the key authors, journals, and themes discussed.

1.3 Key Themes in Analytics Research

As detailed, this section will seek to evaluate the analytics research area. To do so we employ a computational literature review (CLR), presented in Mortenson and Vidgen (2016). The CLR, a quantitative alternative to systemic literature reviews, provides a data-driven analysis of the content and networks within the prior research. Accordingly, it offers not only a convenient and efficient tool to assess prior research, but also is emblematic of some of the opportunity and change proffered by analytics and big data. Accordingly, in the spirit of art-imitating-life (and vice-versa), the tool represents not only suitable research instrument for charting prior literature into analytics, and the significant components within it, but also an exemplar of the sort of change to academic practice the OR discipline could engage with.

The CLR takes an export of abstracts and meta-data related to a search keyword (e.g. "analytics") and provides three primary areas of analysis:

1. Impact: metric analyses of the influence of authors, journals and papers;
2. Structure: analysis of co-authorship using network component analysis;
3. Content: an analysis of the key topics in the abstracts determined using latent Dirichlet allocation (LDA), as first presented in Blei *et al.* (2003). (Note: LDA is also used as a primary method of analysis in this research. Therefore, a description of the approach is given in the research methodology - chapter 2).

For the purposes of this research, only the impact and content elements were deemed relevant. In other words, we seek to identify the most influential papers in this area (by reviewing publication and citation frequencies (and related metrics), the most common themes, before finally evaluating the most influential papers `most relevant to this study (based on filtering by content and sorting by impact). In particular, we seek to explore further the areas that most related to the OR discipline, as this is an appropriate indicator of the intersection of the discipline and analytics in respect to academic research. Each of these elements will be presented in sequence.

1.3.1 Impact

The abstracts and meta-data used in this analysis were extracted from the Scopus database (https://www.scopus.com/), requiring the keyword "analytics" to be included in the article title, abstracts or keywords. The search was limited to academic journal and conference papers in fields relating to computer science, engineering, social sciences, mathematics, business, decision sciences, economics, and health professions and publication dates between 2000 and 2016. After eliminating titles that were duplicated or were missing abstracts, a total dataset size of 9,750 papers was used in the analysis.

The first notable output from the CLR is an analysis of the quantity of publications by year, as presented in figure 3. In line with the earlier analysis of Google searches (figure 1), the research area has significantly grown over this period, although in comparison to overall internet searches there is a far more evident spike from 2010, with the year 2014 boasting 13,833% increase on the 2009 figure (2,490 papers in 2014 in comparison to just 18 in 2009). The dramatic nature of the increase suggests something about the susceptibility of research to "hype cycles", something which be explored further in chapter three.

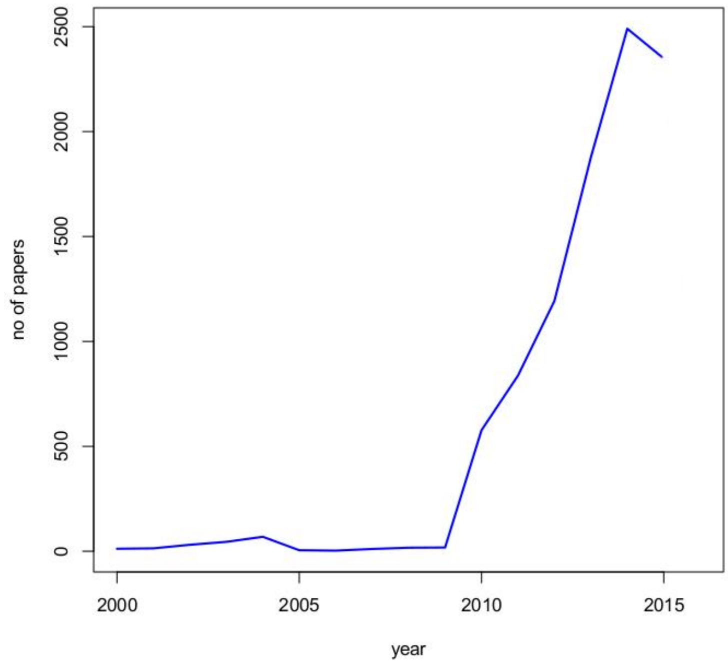


Figure 3 Volume of "analytics" papers published per year (2000-2015)

Secondly, we evaluate the articles which have the highest impact based upon citations. However, to control, to some extent, for recency (on the basis that more recent papers have not had the same opportunity to accumulate citations that older papers have had) we sort the results by averaged citations per year (subtracting from 2016). These results are shown in table 1.

Unsurprisingly, as keyword searches are never perfect, there are a handful of results which are (arguably) unrelated to the research (e.g. Sau *et al*, 2010; Krämer *et al*, 2014; Cho and Patten, 2007; Woodford, 2011; Kloss *et al,* 2012). Evaluating the remainder, the top results, again unsurprisingly are for the more generalist works. Han *et al.* (2012), the most cited, provide a broad and wide-ranging overview of data mining, data management and related techniques. Boyd and Crawford (2012) question the emergence of big data and its potential transformative power to the methodologies of "economic, social, technical, and legal" enquiries. They conclude that whilst there is indeed much that is 'new' and revolutionary in big data analytics, the seeds and the overall concerns of traditional science remain, and that in embracing these approaches a wide range of ethical, political and social concerns need also to be addressed. McAfee and Brynjolfsson (2012), writing in Harvard Business Review, take a management-orientated perspective on big data, considering the potential business value that utilising big data and analytics may bring.

Table 1 Most impactful analytics papers (citations per year)

Rank	Authors	Title	Cites	Cites P.A.
1	Han *et al.* (2012)	Data Mining: Concepts and Techniques	1038	207.60
2	Boyd and Crawford (2012)	Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon	319	63.80
3	Chen *et al.* (2012)	Business intelligence and analytics: From big data to big impact	314	62.80
4	Sau *et al.* (2010)	Properties and applications of colloidal nonspherical noble metal nanoparticles	416	59.43
5	Krämer *et al.* (2014)	Causal analysis approaches in ingenuity pathway analysis	118	39.33
6	Castillo *et al.* (2011)	Information credibility on Twitter	185	30.83
7	Mathioudakis and Koudas (2010)	TwitterMonitor: Trend detection over the twitter stream	184	26.29
8	McAfee and Brynjolfsson (2012)	Big data: the management revolution.	129	25.80
9	Suh *et al.* (2010)	Want to be retweeted? Large scale analytics on factors impacting retweet in twitter network	162	23.14
10	Cho and Patten (2007)	The role of environmental disclosures as tools of legitimacy: A research note	229	22.90
11	Domingos (2012)	A few useful things to know about machine learning	114	22.80
12	Zhang *et al.* (2011)	Data-driven intelligent transportation systems: A survey	135	22.50
13	Buch *et al.* (2011)	A review of computer vision techniques for the analysis of urban traffic	125	20.83
14	Woodford (2011)	Simple analytics of the government expenditure multiplier	110	18.33
15	Kloss *et al.* (2012)	Models, algorithms and validation for opensource DEM and CFD-DEM	87	17.40
16	Andrienko *et al.* (2010)	Space, time and visual analytics	119	17.00
17	Bonomi (2012)	Fog computing and its role in the internet of things	84	16.80
18	Ferguson (2012)	Learning analytics: Drivers, developments and challenges	80	16.00
19	Xu *et al.* (2014)	Internet of things in industries: A survey	46	15.33
20	Kitchin (2014)	The real-time city? Big data and smart urbanism	38	12.67

Alongside the prominence of big data as a topic, which further exemplifies the link with analytics previously discussed, there are several other research areas evident in this list.

Most obviously, these include social media data, specifically Twitter (Castillo *et al,* 2011; Mathioudakis and Koudas, 2010; Suh *et al*, 2010); smart cities (Zhang *et al*, 2011; Buch *et al*, 2011; Kitchin, 2014); and Internet of Things (IoT) research (Bonomi, 2012; Xu *et al*, 2014). The different topics presented in analytics research are discussed further in section 1.3.2.

Secondly, we evaluated the journals that published the highest quantities of analytics research. The top 20 results are presented in table 2, sorted in order of volume of publications, and featuring volume of citations and impact (average citation per paper). Computer science dominates this list, not least with the top two overall venues, which represent 1,069 publications between them (11% of the total number of papers across the whole collection). The information systems (IS) discipline is also present, with three venues in the top 20 and several others further down the list. Only one OR journal makes the top 20, INFORMS' *Interfaces*, with, the more interdisciplinary, *Decision Support Systems* (28th); the *European Journal of Operational Research* (55th); and the *Proceedings of the Winter Simulation Conference* (82nd) also making the top 100 venues.

Table 2 The top 20 journals for analytics research (ranked by frequency)

Rank	Source	Frequency	Cites	Impact
1	Lecture Notes in Computer Science	669	386	0.58
2	ACM International Conference Proceeding Series	400	95	0.24
3	CEUR Workshop Proceedings	171	20	0.12
4	Proceedings of SPIE - The International Society for Optical Engineering	140	92	0.66
5	IEEE Transactions on Visualization and Computer Graphics	126	959	7.61
6	IBM Data Management Magazine	101	2	0.02
7	Proceedings of the VLDB Endowment	100	545	5.45
8	ACM SIGMOD International Conference on Management of Data	90	891	9.90
9	Communications in Computer and Information Science	79	24	0.30
10	Lecture Notes in Business Information Processing	77	48	0.62
11	Annual Hawaii International Conference on System Sciences	73	40	0.55
12	Procedia Computer Science	67	64	0.96
13	Conference on Human Factors in Computing Systems - Proceedings	62	197	3.18
14	2014 IEEE Conference on Visual Analytics Science and Technology	61	0	N/A
15	International Conference on Data Engineering	53	161	3.04
=	International Conference on Information and Knowledge Management	53	160	3.02
17	IEEE Computer Graphics and Applications	49	208	4.24
18	2014 IEEE International Conference on Big Data	48	0	N/A
19	ACM International Conference on Knowledge Discovery and Data Mining	41	236	5.76
20	Information Visualization	40	131	3.28
=	Interfaces	40	64	1.60

Finally, we evaluate the most impactful authors, based upon the widely-used h-index metric. The top 20 authors on this basis are shown in table 3, along with the institute they are currently based at (using Google Scholar data). Perhaps unsurprisingly, all the authors are based in the US (9 in the top 20), Europe (8) and China (3). Evaluating the areas these authors are associated with; the most frequent are visual analytics and machine learning.

Such analyses provide insight into the overall structure of the analytics research area, and the key papers, journals and authors. However, as this analysis and prior discussion would suggest, analytics is a broad church. For many of these metrics to be truly meaningful, such results need to be contrasted with the specific topics and research areas within which they are focused. This will be the concern of the remainder of the section.

Table 3 The top 20 authors in analytics (ranked by h-index and citations)*

Rank	Author	Affiliation	Papers	Cites	Impact	H-index
1	Andrienko G	City University London	40	498	12.45	12
2	Andrienko N	City University London	39	491	12.59	12
3	Keim D	University of Konstanz	15	180	12.00	8
4	Ebert DS	Purdue University	27	166	6.15	8
5	Ribarsky W	University of North Carolina	29	164	5.66	8
6	Dou W	Nanjing University	20	131	6.55	8
7	Liu S	University of Southern California	32	227	7.09	7
8	Schreck T	University of Konstanz	21	202	9.62	7
9	North C	Virginia Tech	29	169	5.83	7
10	Elmqvist N	University of Maryland	21	154	7.33	6
11	Endert A	Georgia Tech	24	146	6.08	6
12	Wang X	Chinese University of Hong Kong	29	113	3.90	6
13	Ertl T	University of Stuttgart	21	113	5.38	6
14	Ma K-L	University of California at Davis	12	98	8.17	6
15	Maciejewski R	Arizona State University	21	92	4.38	6
16	Van den Poel D	Ghent University	9	85	9.44	6
17	Keim DA	University of Konstanz	33	83	2.52	6
18	Chang R	Tufts University	22	83	3.77	6
19	Zhang D	Macau University of Science & Technology	18	66	3.67	6
20	Holzinger A	Graz University of Technology	12	61	5.08	6

* From further investigation "Keim D" and "Keim DA" are in fact the same author. As other authors have not been checked in this way, we continue to treat these as separate.

1.3.2 Content

As discussed, another major component of the CLR is an analysis of the content of the research area; in this case, the specific topics and/or sub-topics prevalent in the articles. This is achieved through the LDA algorithm. As detailed in Mortenson and Vidgen (2016), this stage requires the researcher to input a value for K, the number of topics. In order to maximise the impact of this part of the analysis, several values were tested and the results assessed, before settling on a value of K=45, as this seemingly produces the most meaningful results. (Further discussion on the selection of K is presented in the research methodology as it pertains to the use of these methods in the empirical research).

Having determined this parameter, the final model can be run and the outputs analysed. Following the guidelines in Mortenson and Vidgen (2016), the first step is determining the topic labels; that is the subjects discussed in the abstracts. This is achieved through a visual analysis of the topic clouds generated (included as appendix item A) and, where required, reviewing the papers that load highly on (i.e. have the greatest proportion of their content focused on) the topic. Following these processes, the suggested topic labels were generated, shown in table 4 (along with the number of papers that have this as the main topic).

Table 4 Topics in analytics abstracts (ranked by frequency as main topic)

Topic	Label	Main topic	Topic	Label	Main topic
6	Learning analytics	771 - (0.36%)	4	Statistics	165 - (0.08%)
35	Visualisation	508 - (0.23%)	7	Realtime analytics	159 - (0.07%)
15	Social media	423 - (0.20%)	13	Predicitive models	158 - (0.07%)
28	Business & management	410 - (0.19%)	39	Sports & games	157 - (0.07%)
21	Big data	383 - (0.18%)	20	Knowledge & intelligence	148 - (0.07%)
5	Video & image data	332 - (0.15%)	31	Decision making	137 - (0.06%)
18	High performance computing	325 - (0.15%)	27	Technology	134 - (0.06%)
34	Databases & data management	317 - (0.15%)	33	Process & product design	133 - (0.06%)
22	Big data tools	313 - (0.14%)	14	Literature reviews	128 - (0.06%)
32	Energy	306 - (0.14%)	10	Collaboration	124 - (0.06%)
43	Text data	294 - (0.14%)	26	Questions & objectives	121 - (0.06%)
9	Health	285 - (0.13%)	19	Supply chain & systems management	117 - (0.05%)
40	Politics & culture	280 - (0.13%)	17	Service industries	115 - (0.05%)
45	Smart cities	269 - (0.12%)	29	Data mining	107 - (0.05%)
8	Cloud computing	264 - (0.12%)	11	Psychographics & behaviour	103 - (0.05%)
16	The web	256 - (0.12%)	23	Optimisation	89 - (0.04%)
25	Physical science	250 - (0.12%)	1	Information systems	74 - (0.03%)
3	Security	247 - (0.11%)	37	Information	71 - (0.03%)
24	Mobile & IoT	229 - (0.11%)	36	Volume & scalability	59 - (0.03%)
30	Machine learning	228 - (0.11%)	12	Hiearachies & dimensions	54 - (0.02%)
42	Finance	219 - (0.10%)	44	Frameworks	54 - (0.02%)
38	Science & academia	217 - (0.10%)	41	Importance & growth	37 - (0.02%)
2	Software	180 - (0.08%)			

As can be seen, topics vary across a range of 'types'. There include:

- Disciplines (such as "statistics" or "information systems");
- Methods, practices or applications (e.g. "visualisation" or "optimisation");
- Domains (e.g. "physical science" or "finance");
- Research practices (e.g. "literature reviews" or "frameworks").

The last of these types (research practices) is less relevant to our specific study, although obviously this may differ in other projects. However, across the remainder there are clear links to the earlier discussion. Most obviously is the prominence of big data related topics (e.g. "big data", "big data tools" and "volume & scalability"), further supporting association with analytics. As with the top papers analysis, there are topics relating to visual analytics, social media, IoT and smart cities. Perhaps surprisingly, the most frequent 'main topic' is "learning analytics". At first glance, this topic may seem highly relevant, but, moreover, these papers relate to the use of analytics to support education and student engagement (the papers that most highly load on this topic are Tempelaar *et al.* (2014) and Lonn *et al.* (2015), which focus on these subjects) rather than the teaching of analytics. In other words, "learning" in this context is a part of a bigram, not a verb. Whilst unexpected, the prominence of this topic can be relatively easily explained. Ultimately most researchers are also teachers, and therefore the education domain is clearly relevant.

Thereafter, the next highly ranked topics (in terms of the main topic metric) are more expected. "Visualisation", "high performance computing" and "databases and data management" are clearly highly relevant, whilst "social media" and "video and image data" represent obvious choices of data source in analytics, if we accept the correlation it has with big data. In respect to OR, there are many indirectly related topics, including many of the domains topics, and two more directly related topics in "optimisation" and "predictive models" (which features "simulation" as a prominent term in the topic cloud). For this reason, we investigate these two directly related topics further, as well as "business and management" (the most generic domains topics). These topics will be analysed in sequence.

1.3.2.1 The 'Optimisation' Topic

The optimisation topic is the most clearly affiliated to OR. Its label was reached from an analysis of the topic cloud, shown as figure 4. As can be seen, many of the key terms (each word is sized in scale with its frequency in the topic) have clear associations with optimisation (such as "optimization"/"optimal"; "problem", "constraints", and "solving").

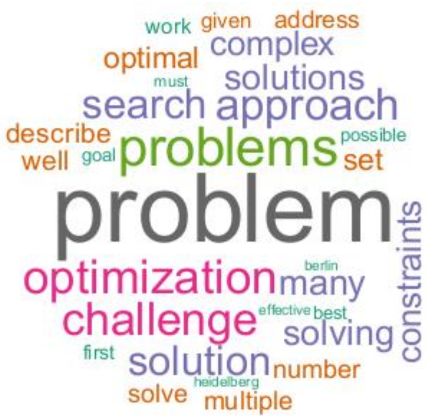


Figure 4 The 'optimisation' topic cloud

Whilst the quantity of papers featuring this as their 'main' topic is not particularly high (0.04% - 39th overall), there could be good reason for this. Many papers using optimisation methods will not necessarily be about optimisation *per se*, a reasonable assumption could be made that optimisation represents effectively a means to an end, with the domain (or similar) recorded as the main topic – i.e. the subject that is discussed most in the paper. As a relevant topic, using the derived scores for each document related to the topic (effectively the proportion of the abstract discussing optimisation) we created a subset of documents by filtering to incorporate only abstracts that have at least 10% of content associated with this topic. This subset is then ranked by citations, effectively identifying the most impactful papers associated with the topic. The top 10, based on this criterion, are shown in table 5.

Table 5 The top 10 papers associated with the ‘optimisation’ topic

Authors	Cites P.A.	Description
Kasprzyk *et al.* (2013)	10.75	Combining evoloutionary opimisation and robust decision making techniques, the paper presents an approach for managing complex environmental systems. An additional "analytics" element is the use of visual analytics tools to make selections from the solution space.
Kasprzyk *et al.* (2012)	7.40	Using a "risk-based water supply management problem", the paper demonstrates a framework based (again) on evoloutionary optimisation. Aside from the algorithm presented itself, there is no additional "analytics" content.
Woodruff *et al.* (2013)	6.50	The paper presents the "MOVA" framework which provides a visual analytics tool for comparing solutions to optimisation models, and allowing the researcher to vary the number of objectives.
Fu *et al.* (2013)	4.00	Based upon the "Anytown" network problem, the authors (again) present a solution that combine evolutionary optimisation and visual analytics, to evaluate trade-offs.
Liu *et al.* (2009)	3.88	Detailing research that seeks to present text summaries (using LDA) and interactive visual analytics tools to allow the researcher to more easily interperet complex and dense results.
Isenberg *et al.* (2012)	3.40	Describes a software tool designed to provide teams with a visual analytics tool to (manually) solve a multi-criteria decision problem. The study also incorporates aspects of decision science to analyse results.
Mutschler *et al.* (2013)	2.75	Details a competition based on a sports analytics case 'problem'. Partcipants had to query real-time data, in a distributed architecture setting, to find the optimal solution.
Giuliani *et al.* (2014)	2.33	The research describes an extension to Multiobjective Markov Decision Processes optimisation to allow for more objectives. This is achieved through non-negative principal component analysis.
Mian *et al.* (2013)	2.25	The goal of the work is to provide solutions to the multi-objective problems for delivering optimal configuration for data analytics in a cloud environment. Two principal objectives are resource costs, and SLA penalties.
Tauer *et al.* (2013)	2.00	Descibes a solution to the graph assosciation problem - whereby multiple graphs with potentially shared relationships are combined. The authors develop a Lagrangian heuristic which outperforms CPLEX on large problems.

The first observation is that not all of these papers describe optimisation in the traditional, mathematical sense of the word. However, in papers such as Liu *et al.* (2009) and Isenberg *et al.* (2012), the goals remain aligned with optimisation, even if the methods are not; both seek to identify ‘optimal’ solutions to multi-criteria problems. Accordingly, the association with the topic identified by the model does seem to hold. Secondly, it is clear that many of the papers listed are exploring similar territory (indeed many feature the same co-authors); namely the use of visual analytics and multi-objective optimisation in the environmental domain. In some cases, the “analytics” component of the study is only in the optimisation methods used. Depending on how one chooses to partition analytics and OR, a case can be made that these papers are affiliated with the latter and not the former. These issues, in combination with the relatively low number of articles featuring “optimisation” as their main topic, suggests a potentially limited role for OR within prior analytics research.

1.3.2.2 The 'Predictive Models' Topic

As before, we begin the analysis by reviewing the topic cloud generated in the CLR (figure 5). From evaluating the most frequent terms, several can be associated with the label given ("predictive", "accuracy" and "models"), as well as those associated with OR (albeit not exclusively), such as "simulation", "forecasting", "regression" and "Bayesian".

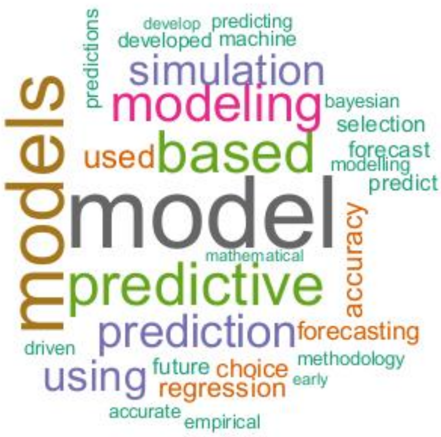


Figure 5 The 'predictive model' topic cloud

Again, the top 10 papers were analysed as shown in table 6. In terms of domains in which they are focused upon, a slightly wider spread is evident, incorporating credit, marketing, healthcare and information systems research. In respect to analytical techniques, perhaps unsurprisingly, predictive analytics are the mainstay, in particular machine learning. As to the prominence of OR, there is less evidence than may be expected. Di Domenica *et al.* (2007) incorporate simulation, stochastic modelling and decision analysis, but, whilst machine learning and statistical methods relevant approaches used in OR research, no other papers in this subset are particularly aligned with the OR discipline and methodology.

1.3.2.3 The 'Business and Management' Topic

The final topic considered is of 'business and management'. This selection is based in part upon a general alignment of OR to the domain. Indeed, most UK academic OR groups are based in business schools (evidenced by 11 out of the 18 OR research groups identified by the OR Society (www.theorsociety.com/Pages/Research/ResearchWhoDoes.aspx) being primarily based in business schools). However, additionally the topic features highly in terms of main topic rankings (4th) and also has a general alignment to the overall project. As with the previous analyses, we begin by presenting the topic word cloud (figure 6).

Table 6 The top 10 papers associated with the 'predictive models' topic

Authors	Cites P.A.	Description
Shmueli *et al.* (2011)	12.50	Exemplifies some use cases of predictive analytics as a tool for IS research, and discusses the potential theoretical and practical contributions the use of such tools could have.
Armstrong *et al.* (2015)	6.00	Discusses the virtues of conservatism in producing forecasts. Gives a fairly in-depth explanation of key components of such forecasts.
Khandani *et al.* (2010)	3.57	Describes the use of a (non-parametric) machine learning approach to a forecasting task in credit risk. Their models show performance improvement over traditional approaches, and is estimated as being more robust in the face of systemic risks.
Miguéis *et al.* (2012)	3.20	The authors introduce a predictive analytics model to forecast potential customer churn in a grocery store setting.
Di Domenica *et al.* (2007)	2.60	The authors describe their "multifaceted view of modelling", taking in stochastic resource allocation, simulation and post-modeling data analyses, which they describe as, in combination, "business analytics". They also discuss the integration of such approaches with OLAP and decision support systems.
D'Haen *et al.* (2013)	2.50	Describing ensemble machine learning approaches to predicting customer profitability, the article presents an effective model to this end. Also evaluated is the data sources used, comparing web data and commercially acquired data (with the former outperforming the latter).
Ng *et al.* (2014)	2.33	The paper presents a software tool designed to facilitate and ease the processes required for performing predictive modeling on healthcare date.
Ballings *et al.* (2012)	2.00	The author's evaluate the extent of customer history should be recorded to effectively predict customer churn. Using a combination of logistic regression and bagging, the results suggest that after 5 years only marginal gains are seen.
Krause *et al.* (2014)	2.00	Presenting a visualisation tool that supports predictive analytics on high-dimensional datasets by assisting with feature selection.
Shin *et al.* (2014)	2.00	A 'proof-of-concept' model is discussed, designed for the processing and analysis of Big Data in a manufacturing setting (specifically the metal cutting industry). The tool is based upon the Hadoop Distributed File System (HDFS).

The word cloud demonstrates, relatively clearly, the appropriateness of the given label, with prominent terms such as "business", "management", "customer", "companies" and "organizations". This is also evident in many of the highest ranked papers (table 7), covering areas such as business intelligence, process management and human resources.

Figure 6 The 'business and management' topic cloud

Table 7 The top 10 papers associated with the 'business and management' topic

Authors	Cites P.A.	Description
Elbashir *et al.* (2011)	7.17	The paper evaluates the conditions that determine the success (or otherwise) of business intelligence and analytics tools. Their findings point to "absorptive capacity" as a critical factor.
Bijmolt *et al.* (2010)	6.71	The authors discuss the 'state of the art' in respect to analytical models for customer engagement. Thereafter, the paper focuses on how these align to consumer behavioural models, and discuss organisational implications.
Barton and Court (2012)	5.60	The article discusses the growth of analytics and Big Data, but primarily from an management perspective. The authors detail best practice in data management and acquisition, model selection and implementation, as well as embedding analytics in the organisation.
Thorleuchter *et al.* (2012)	5.40	Details the use of text analytics and predictive models to better understand prospective B2B customers, to inform acquisition and marketing activity.
Wixom *et al.* (2014)	5.00	The article is an analysis, mostly via a questionnaire, of business intelligence and analytics provisions in universities. This includes a summary of covered modules, as well as discussion of industry requirements.
Vera-Baquero *et al.* (2013)	4.75	Presents an architecture for managing, mining and analysing the output data of large and complex supply chains from the perspective of process monitoring and optimisation.
Janiesch *et al.* (2012)	4.40	Describes an architecture to use in process architecture which can allow for real-time management of business processes. The architecture produces events (simulated), processes events (both simulated and real), and consumes events (creating visualisations, alerts and/or automating decisions).
Abrahams *et al.* (2013)	4.25	Presents a solution for businesses to identify business insights from the 'noise' of social media. The model is demonstrated on an automotive repairs scenario.
Aral *et al.* (2012)	4.20	The authors describe the benefits of integrating IT, performance-related pay and HR analytics, demonstrated using a principal agent model.
Phippen *et al.* (2004)	3.92	The article describes the rise and methods of web analytics, in particularly with reference to how the insights generated can be used to shape business strategy.

One potentially surprising result is the inclusion of Wixom *et al.* (2014), as the paper is primarily concerned with the development of curricula in universities. However, it is important to note that the top 10 is derived based on citations (and a minimum proportion of content related to the topic), so this is not a measure of the proportion of the paper dedicated to the topic. It is also noteworthy that actually "business and management" is jointly its "main topic", alongside "science and academia" (13.22% in each case). Finally, in presenting much content on the requirements of businesses for graduates, the centrality of this topic does make some sense.

Evaluating the papers from an OR perspective, few directly describe the methods and applications most associated with OR. Janiesch *et al.* (2012) is somewhat the exception to this, however, the paper is perhaps better described as "interdisciplinary" rather than solely within the OR tradition.

1.3.3 OR, Analytics and the Publishing Paradox

The discussion of this section into analytics research has highlighted many of the key concerns and reaffirmed the correlation between the subject and big data (as well as other topics). However, it also has suggested that, despite the potential benefits of an association with analytics may have (e.g. Liberatore and Luo, 2010), there is evidence that analytics research from within the OR research community may be somewhat lacking. To further demonstrate this, we evaluate the number of publications in our collection (9,750 articles) which originate from OR journals. We draw the list of such journals from the Association of Business Schools’ 2015 *Academic Journal Guide* (http://charteredabs.org/academic-journal-guide-2015/), which lists 64 OR-related journals, which, although clearly not exhaustive (particularly as this excludes conference papers), serves as a suitable sample.

Using this sample to count the numbers of publications, only 13 in total (20.31%) feature in this collection (inferring only 13 have published analytics content), the list of which are shown in table 8. This relates to an average of 1.39 papers per journal, and less than 1% (0.91%) of the total research output examined in this study (the 9,750 articles extracted).

Table 8 The number of ‘analytics’ publications in OR journals

Journal	Pubs	Journal	Pubs
Interfaces	40	Reliability Engineering and System Safety	2
European Journal of Operational Research	16	IEEE Transactions on Systems, Man, and Cybernetics: Systems	1
Management Science	9	Journal of Heuristics	1
Annals of Operations Research	9	Naval Research Logistics	1
Decision Analysis	3	OR Spectrum	1
International Transactions in Operational Research	3	Asia-Pacific Journal of Operational Research	1
Journal of the Operational Research Society	2		

From further analysis, of the 89 published, nearly half (44.94%) have been published in just one venue, INFORMS’ *Interfaces*. If this journal were excluded then OR journals would account for only 0.50% of total publications, and would average 0.77 papers per journal.

This would suggest a clear discrepancy between the perceived benefits that an association with analytics may offer the OR discipline (e.g. Ranyard *et al*, 2015) and the amount of research into analytics emanating from the OR research community. There are of course many limitations in this approach, as (1) is dependent on the database returning all results; and (2) ignores the possibility that OR researchers are publishing analytics-related content in non-OR journals (which is of course highly plausible). Whilst these results cannot be considered indubitable, they still act as a strong indicator that there is a disparity between the perceived value of OR research into analytics and the volume of such research.

This is somewhat tempered by an increased output observed since 2012, as shown in figure 7 which lists analytics publications per year (up to March 2016) across OR journals, and the recent release of Decision Analytics in 2014, a journal that features both analytics and OR content. However, considering that the first academic articles discussing analytics were published in the early 2000s (e.g. Kohavi *et al*, 2002), the tardiness of the OR academic community's response is surprising enough to warrant further exploration of the causes.

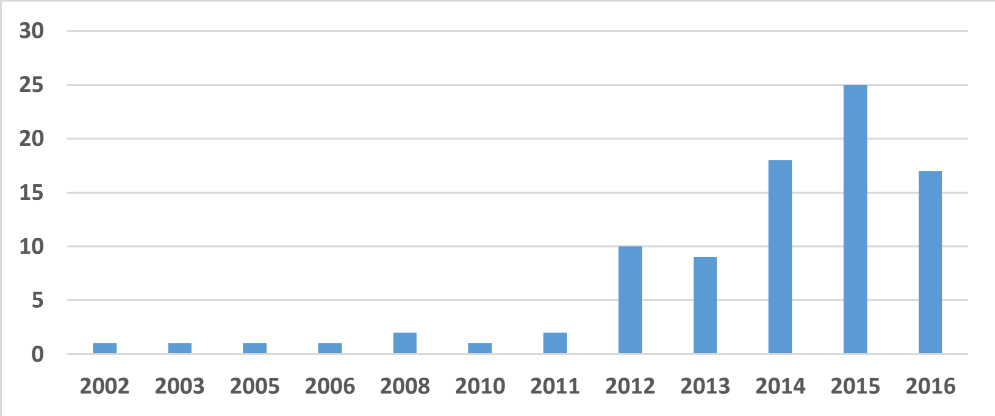


Figure 7 Number of 'analytics' publications per year in OR journals

Although primarily this study is concerned with university *teaching*, obviously research is a key part of the business of 'UK universities' (as indeed it is worldwide). Whilst we primarily seek to assess the degrees and modules related to OR and analytics, and the extent to which they prepare graduates for future work in these areas, it would be remiss to not also consider research efforts in these areas. Additionally, the results of the CLR clearly suggest this to be a gap, and that there is a relative parity of OR research into analytics, meaning such efforts offer a potentially valuable contribution. Accordingly, we posit the following as the second research objective of this study:

RO2: To develop a research agenda for the OR community which addresses the concerns associated with analytics.

This section has discussed an overview of the analytics literature and, in particular the role which OR research plays within it. To return more directly to the specifics of the research question, the next section reviews literature related to analytics employment and job roles.

1.4 Analytics and OR: Job roles and responsibilities

Although much of the literature on analytics is focused on new stimulus and new challenges in modern business (as briefly discussed in section 1.1), analytics as a profession, if indeed it is one, cannot simply emerge from no-where.

In *The Systems of Professions*, Andrew Abbot (1988) addresses how "expert" professions emerge and are demarcated. He characterises this process as more multi-directional and amorphous than previous studies had treated them. In other words, professions, or moreover the collection of tasks, activities and competencies to which its practitioners claim expertise, are not necessarily 'set in stone', and can evolve and update over time.

Expanding on these ideas, with specific reference the OR profession, Corbett and van Wassenhove (1993) discuss a "drift" from the initial actualisation of the discipline as a form of "management engineering" (adapting analytical methods to meet specific problems), into both "management science" (the development of methods and toolkits) and "management consulting" (the deployment of analytic methods to specific problems, but in a more 'verbatim' way than management engineering). Considering this example in respect to the concept of an analytics profession, and particularly from where it has drawn its 'professionals', there can be several inferred pathways, from a more client-orientation (management consulting), which we may infer prioritises an ability to evaluate management problems and prescribe a suitable method; through to a scientific-orientation that may be closer to the hard sciences, and the technical development of methodologies and tools.

Additionally, however, Abbot points to some 'competition' between professions as to 'ownership' of these expertise. Such consideration seems highly relevant to discussion of an "analytics profession", particularly in respect to what may be demarcated as its specific territory, and in how it may be differentiated from the work of an OR specialist or a computer scientist.

These themes will be the main focus of this section, which addresses the skills requirements for analytics professionals, and the domains in which analytics is applied, having firstly, reviewed the literature comparing OR and analytics in this specific regard.

1.4.1 OR and the Analytics Skillset

As previously stated, there has been effort from some in the OR community to champion the affiliations and synergies between their discipline and analytics. Not least this is evidenced in the activities of INFORMS, the US operational research society. Their analytics offerings include conferences, a dedicated publication (www.analytics-magazine.org) and, most recently, the introduction of a specific analytics certification (https://www.certifiedanalytics.org/). To design this certification, a working group of experts investigated analytics job tasks in order to design their Certified Analytics Professional award, identifying six categories of 'tasks' involved in analytics practice, shown in table 9. In the main, the different tasks listed are likely to be familiar to almost all OR practitioners, and indeed closely resembles many descriptions of OR projects (e.g. Ackoff, 1956; Winston, 2004, pp 1-7). As a note, a full description of the OR profession is not included, as it is believed to be outside of the scope of this chapter and has been covered in detail elsewhere in the literature. The sources above are a good reference point for the interested reader, as well as several others (e.g. Churchman, 1970; Corbett and van Wassenhove, 1993; Ranyard *et al*, 2015)

Table 9 Analytics 'domains' included in INFORMS' CAP program

(Source: INFORMS, 2014)

Domain	Approximate Weight
I. Business Problem (Question) Framing	12%–18%
II. Analytics Problem Framing	14%–20%
III. Data	18%–26%
IV. Methodology (Approach) Selection	12%–18%
V. Model Building	13%–19%
VI. Deployment	7%–11%
VII. Model Life Cycle Management	4%–8%

One potential critique of this, is that essentially it represents an OR group's interpretation of analytics domains, which may not be the same perspective those from other communities would bring. Many other authors discuss alternative approaches to analytics, where a combination of big data and machine learning / data science methods are used to identify interesting correlations without the need to build particularly extensive models, nor explicitly test a priori hypotheses (e.g. Anderson, 2008; Mayer-Schönberger and Cukier, 2013). Well known examples of this include the Google Flu Trends engine (e.g. Carneiro and Mylonakis, 2009) and the Netflix prize competition for recommending online videos. Another possible issue is the potential for sample bias as of the eleven members of the initial expert panel determining the list, only two were not already INFORMS members (INFORMS, 2014).

Another related study, published in INFORMS' *Interfaces* journal, is given in Liberatore and Luo (2013). Their research is centred on the results of a questionnaire of (effectively) 1,206 INFORMS members and readers of Analytics magazine which asked respondents to state the importance of hard and soft skills associated with both professions (ranking from the same two lists (hard and soft) once for analytics and once for OR). Their findings show some differences, such as the relative prominence of data management and "business-orientated skills" (ibid, p 197).

A similar study forms part of paper presented by Ranyard *et al.* (2015), although this time focused on a more international sample, members of the International Federation of Operational Research Societies (IFORS). Again, the principal instrument is a questionnaire, although the scope is wider reaching in the sense that it considers multiple aspects of OR practice rather than a comparison with analytics practice (the questionnaire is available at: https://sakai.lancs.ac.uk/access/content/user/hut2/IFORS%202013/IFORS%20main%20survey%20launch%202013.03.25.pdf).

Although not explicitly designed to investigate analytics practice, the authors do provide some analysis of this area, by asking participants to rank their familiarity with a range of techniques, and then reducing these using exploratory factor analysis. This analysis finds six components, to which the authors attribute two of which to the practice of business analytics ("Data Mining, Statistics (basic and advanced)" and "Revenue Management, Forecasting and Financial Analysis"; Raynard *et al*, 2015, p9). Using this, a comparison is made with "traditional OR", what may be described as hard OR techniques, and "Decision support/PSMs", the 'softer' OR techniques such as problem structuring methods (PSMs) and strategy generation techniques (*ibid,* p9). The comparison finds that those specialising in analytics were the smallest group, and predominantly based in North America.

There are, however, limitations with both studies. Firstly, there is a danger of sample bias as respondents are likely to have associations with the OR discipline. Indeed, a pre-requisite in the INFORMS survey was that all respondents had familiarity with both analytics and OR roles, whilst Raynard *et al.* (2015) acknowledge that their sample would have obvious bias to OR practitioners. The 'version' of analytics may be very different if the respondents were those familiar with both analytics and, for instance, computer science. Secondly, as Liberatore and Luo (2013) acknowledge, the questionnaire instrument necessitates that relatively few skills can be included in the analysis. Particularly considering these issues in combination, there is the potential that many important skills are being excluded.

Finally, a perspective presented on the OR Society website, INFORMS' UK counterparts, seeks to chart the relationship between the two based upon a division of 'hard' and 'soft' skills (shown in figure 8). The implication of such a separation is that it is just at the 'hard' end of OR that an overlap with analytics occurs. There are issues with the representation. For instance, techniques such as "SWOT" and "PESTLE" are fairly generic approaches certainly not limited to the OR discipline, whilst "stakeholder management" would surely be a task performed by many different individuals, in analytical professions and beyond. This also, indirectly, calls into question the differing role that 'softer' skills may play in OR in comparison to analytics. One possible reading would be that this is more important to the former than the latter. Despite such issues, it does suggest an alternative approach to tackling this issue. Namely, this is through evaluating skills not solely at a 'macro', overall level, but also evaluating them in terms of specific skills, and the categories of technical (hard) and people (soft) skills. This will be the approach adopted in the next section.

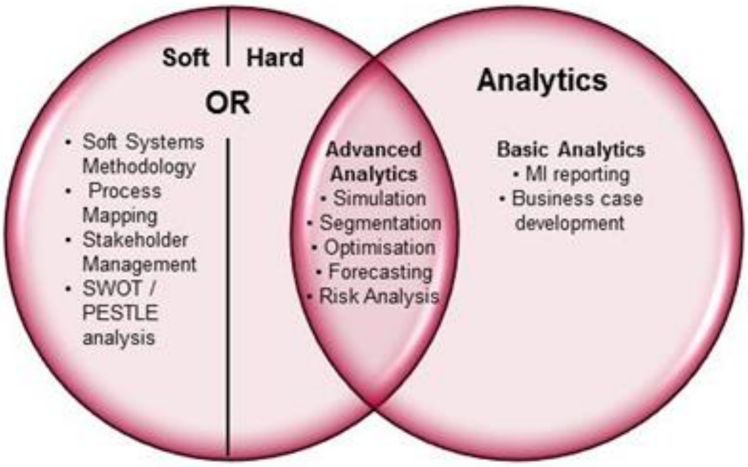


Figure 8 OR, analytics and advanced analytics (Source: Blackett, 2010)

1.4.2 The Analytics Skillset: Hard, soft and MAD skills

The above discussion demonstrates the associations that OR and analytics may have, and therefore makes inference on the skills they may have in common. However, as has already been suggested, a major critique of this work is that such discussion almost unilaterally emerges from the OR community itself. Whilst that certainly does not necessarily disprove the cogency of the argument, it should be tempered by the fact that similar arguments have been made for the role that other disciplines might play, such as information systems (e.g. Chen *et al*, 2012; Molluzzo *et al*, 2015) and computer science (Davenport and Patil, 2012; Mahadev and Wurst, 2015). For example, a recent editorial from the Information Systems Research journal posed the question: "what are the strengths that the information systems (IS) community brings to the discourse on business analytics?" (Agarwal and Dhar, 2014). Therefore, to give a more balanced view on the skills required in analytics, this section will review discussion in this vein from outside the OR tradition.

One such perspective is given in Cohen *et al.* (2009). They argue that the key skills for analytics fall into three categories, using the mnemonic "MAD" skills:

- **Magnetic**: designing and managing data warehouses which "attract" different forms of data from a variety of sources, rather than enforce rigid structures which "repel" new data types;
- **Agile**: working fast and efficiently, named in reference to the software methodology; and
- **Deep**: working with increasingly sophisticated and advanced algorithms and techniques.

OR clearly incorporates the 'deep' part of the mnemonic, and, as most practitioners of the discipline will be regularly involved in project work, has strong associations with project management (e.g. Tavares, 2002), even if this is not necessarily to follow an exact interpretation of the agile framework. However, whilst OR practitioners are heavy users of data, and therefore are regularly analysing datasets from within, or at least extracts from, enterprise data warehouses (EDW), the discipline is less concerned with the design and management of these systems. Indeed, it could be argued that OR has a far more 'ad-hoc' approach to data than disciplines such as data mining (albeit the two have many synergies – e.g. Olafsson *et al*, 2008; Corne *et al*, 2012), which is therefore more closely associated with analyses of varied data types and the use of EDW.

Another potential problem that can prevent an acceptance of the OR version of this relationship, is analytics' association with the much-hyped role of the 'data scientist'. In their initial description, Davenport and Patil (2012) describe the data scientist as a "professional with the training and curiosity to make discoveries in the world of big data". Although there is, as yet, limited academic literature on the role of data scientists, there is a considerable amount in the practitioner literature. Jeff Hammerbacher (2009, p. 84), in a noted example, suggests on a typical day a data scientist at Facebook may:

> "Author a multistage processing pipeline in Python, design a hypothesis test, perform a regression analysis over data samples with R, design and implement an algorithm for some data-intensive product or service in Hadoop, or communicate the results of our analyses [...] in a clear and concise fashion".

Presumably most OR professionals will be more than familiar with many of these tasks, particularly the use of statistical methods and the communication of results. However, authoring a "multistage processing pipeline" or designing algorithms for Hadoop may be less common, certainly in terms of what would be taught on a typical OR degree program. Equally an emphasis on languages such as Python and R may not be obviously associated with the discipline. Further to this, Davenport and Patil (2012) argue, albeit without empirical justification, that the "data scientists' most basic, universal skill is the ability to write code", and whilst some coding is likely in many practical applications of OR, few would argue it to be *the* fundamental skill in the discipline.

Further to this more general discussion, other authors have sought to be more comprehensive in their listings of the skills needed for analytics. A summary of some of these suggestions is shown in table 10.

Table 10 Required skills for analytics and data science

Study	Hard skills	Soft skills
Chiang *et al.* (2012)	Association rule mining; classification; clustering; neural networks; deviational analysis/anomaly detection; geo-spatial and temporal analysis; network/graph analysis; sentiment analysis; optimization; simulation; decision trees; logistic regression; forecasting; time series; relational databases; data warehousing; ETL/OLAP; visualisation; dashboard design; text/Web mining; massive data file systems (Hadoop); MapReduce; unstructured data management; social media; web services; APIs; search engines; cloud computing; mobile.	"Understanding business issues and framing the appropriate analytical solutions. Listening to what the business needs and being aware of what the business intends to accomplish is of fundamental importance. At a minimum, the necessary business domain knowledge […] includes fundamental knowledge in the areas of accounting, finance marketing, logistics, and operations management"
Laney & Kart (2012)	Big data; machine learning; computing; algorithms; programming (SQL, R, C, SAS, Python, Java, Hadoop and Pig); data management (integration, manipulation, quality assurance, preparation); analytics modelling (techniques, interpretation and model diagnostics).	Team work; communication skills; business analysis (goals, constraints and decisions); collaboration; leadership; creativity; discipline; passion.
Davenport & Patil (2012)	"Bring structure to large quantities of formless data"; data joining; data cleaning; data hacking; data analysis; programming; software development; academic research; experimental design.	Creativity; visualisation; communication; story telling; curiosity.
Dhar (2013)	Machine learning; statistics; econometrics; data structures; algorithms; database skills; programming (e.g. Python and Perl); Hadoop.	Problem formulation.
Sanders (2013)	Hadoop; Java; Python; SQL; Hive; Pig; ETL; fact tables; data warehouses; R; Excel; SAS; statistics; mathematics; Tableau.	Business domain expertise; presentation skills; storytelling skills; PowerPoint
Manyika *et* al, (2011)	A/B testing; association rule learning; classification; clustering; data integration; signal processing; natural language processing; data mining; genetic algorithms; machine learning; neural networks; network analysis; optimisation; sentiment analysis; spatial analysis; statistics; simulation; time series.	Visualisation

The results of table 10 again show indication of the role of computing technologies and skill. All bar the last (Manyika *et al*, 2011) discuss programming as whole and/or specific languages. All of them, to some extent, discuss data management tasks alongside data analysis and modelling. Whilst some might argue that “data science” and “analytics” are distinct fields (the majority of skills listed in table 10 refer to the former), and that “data science” may have more of a technology orientation than “analytics”, there is little in the literature supporting this view, or from which to draw particularly clear boundaries in this regard. The role of “data scientist” may be a little more separable, though with a typical description as a those who “understand analytics, but they also are well versed in IT” (Davenport *et al*, 2012), the implication is that there is significant communality. In many ways, the role of data scientist can simply be read as a professional versed in the “full stack” of the analytics process.

As may be expected, the listings for soft skills are a little more ambiguous. Equally, as with the previous discussion, there is no clear consensus on the relative importance of each, nor, at a higher level, the relative importance of soft skills and hard. Dhar (2013) argues for problem structuring to be the most important of these, a positive indicator for the OR discipline considering its depth of the work in this area. Other areas of importance are seen to be around teamwork, with Laney and Kart (2012) arguing "data scientists are expected to work more within teams than statisticians", and visual, verbal and written communication skills. Finally, many point to the importance of domain knowledge, a subject further explored in the final part of this section.

1.4.3 The Domains in Demand for Analytics Professionals

The previous section has detailed discussion of the more general analytical and technology-orientated skills required of the analytics professional. The results are at times contradictory, and do little to find a clear consensus, a gap to which this research aims to contribute. However, it is likely many jobs in analytics will also have a requirement for skills and knowledge related to the specific domain or industry the role is situated within. Indeed, much of the academic research into analytics (e.g. Kohavi *et al*, 2003) cites the spread of analytical approaches into new domains as one of the main drivers for the growing interest in analytics. Laursen and Thorlund (2010, p xviii) argue analytics should be a concern for "everyone in business-focused functions in sales, marketing, finance, management, production, and HR who works at a strategic level". Further, Pearson and Wagner (2013, p 1) argue that analytics in organisations is so diversified and important to organisations, such to require representation amongst senior management, stating: "if you don't know who (and where) your chief analytics officer is, you may already be behind the curve".

In terms of specifics areas of application, Chen *et al.* (2012, p 1173) list five, "E-Commerce and Market Intelligence; E-Government and Politics; Science & Technology; Smart Health and Wellbeing; Security and Public Safety", albeit mostly as examples rather than an exhaustive list. Davenport and Harris (2007, p 7) highlight a list of "analytic competitors" (companies that are extensive users of analytics in their processes and decision making), sub-divided into the following industry categories: "consumer products; financial services; hospitality and entertainment; industrial products; pharmaceuticals; retail; telecommunications; transport; eCommerce". Additionally, to such examples, more specific literature searches, using "analytics" as a keyword alongside other domains and disciplines reveals a very broad range of work, far too extensive to represent in this part of the study. An example of just some of these is given in table 11.

1.4.4 Summary

Whilst this section demonstrates some clear patterns, there are also many contradictions and open questions. Significantly, there is no consensus as to the specific skills required for analytics roles, and secondly the extent of an overlap with the skills required of OR professionals, again is not comprehensively resolved in the literature.

Table 11 Examples of analytics applications in different domains

Function	Example concerns	Example sources
Marketing	Performance management; market research; customer relationship management (CRM)	Germann *et al.* (2013); Hauser (2007); Nair *et al.* (2013); Peterson *et al.* (2005)
Web	Product recommendation; user segmentation; website design	Chaffey and Patron (2012); Eirinaki and Vazirgiannis (2003); Phippen et al. (2004)
Sales	Performance measurement; CRM	Baier *et al.* (2012); Tanner Jr *et al.* (2005)
Financial	Compliance; risk management; portfolio management	Agrawal *et al.* (2006); Mun (2010; Tezuka *et al.* (2005)
Human resources	Recruitment; staff retention	Kapoor (2010); Pease *et al*. (2012)
Supply chain	Logistics; procurement; product trends	Sahay and Ranjan (2008); Trkman *et al.* (2010)
Strategy	Pricing strategies; strategic planning	Klatt *et al*. (2011); Metters *et al*. (2010)

Prior research, particularly in the form of the questionnaire of Liberatore and Luo (2013), has sought to evaluate some of these issues. However, there are limitations that can be associated with the research instrument in this study, which motivates further investigation of these issues. Therefore, we seek a different approach. As the primary goal is to better understand the skills requirements of analytics and OR professionals, or, in the other words, the typical content of jobs in these fields, a logical source of data would be job adverts. Segmenting and mining these can provide insights into the general trends in the analytics job market, what skills and techniques are most in demand, and ultimately allow us to make inferences as to how analytics is engaged in as a business practice. In doing so, we eliminate some of the risks of selection bias that may be present in a questionnaire instrument.

These methods will be utilised to seek to meet the third research objective of this study:

RO3: To determine the skills requirements of analytics roles and the extent to which these may be met by OR professionals.

Such an objective is not only key to our overall goal, effectively representing one of the key measures to which we can evaluate OR academic provisions, but also is seemingly a further gap (or at least a point of contention) in the current literature. It need be noted that jobs markets are by nature very volatile in respect to skills demands, and that this may be increasingly so in a relatively new, relatively 'hyped', and likely fluctuating area such as analytics. For clarity, this objective can only be held to be relevant to this single moment in time, without guarantees that such skills requirements are necessarily cemented or immovable.

To this end, however, the analysis of the literature presented in this section gives insights into how the analysis may be directed. In particular, it suggests three main areas of investigation that may be utilised:

1. **Hard skills:** The first of these categories would include the more technical elements of analytics such as the computational and quantitative elements of analytics. Whilst most of these skills are of interest, in particular programming languages can be additionally explored, as they have seemingly been identified as particularly relevant in this literature review, and are also comparatively easy to measure (as the names of the languages are relatively unambiguous whereas terms such as "optimise" can have multiple meanings).
2. **Soft skills:** Although, as indicated in table 10, soft skills can be more ambiguous than their technical counterparts, this is also likely to be important to understanding analytics roles. Also noteworthy are the inferences of figure 8 that analytics and OR only overlap in terms of hard skills, and the differences found between rankings of soft skills associated with OR and analytics roles identified in Liberatore and Luo (2013), particularly in the importance of "business-orientated skills" in the latter over the former. In consideration of this, comparing soft skills required in each may provide insight into potential differences between analytics and OR, roles as well as better illustrate overall skills requirements.
3. **Domains:** Again, indicated in the above literature review, one element of analytics growth is seemingly an increasingly wide range of domain applications. As such, it may be hypothesised that analytics and OR domains will differ, with the former potentially reaching a wider range of domains and business functions.

Such a split suggests not only 'areas of interest', but also provides direction for empirical investigation (presented in chapter four). In other words, these areas act as 'performance indicators' for both the types of skill involved in analytics roles, as well as the potential overlap with those of OR professionals.

1.5 Analytics and OR: Courses and education

Having assessed some of the literature regarding the skills required in analytics job roles, the obvious next direction is that regarding what is being taught in our universities. This represents the focus of this section.

Many universities have sought to meet the perceived need for analytics professionals through specialised degrees, with titles such as "Business Analytics" or "Data Science", which have proliferated throughout Europe, North America and the rest of the world. A recent report from Deloitte identifies over 100 analytics-related degree programs in the US alone (Danson *et al*, 2016). Indeed, the value of these ventures has been acknowledged all the way to the governments of such countries. In the forward to a UK Government report, David Willetts and Matthew Hancock (then Ministers for Universities & Science, and Skills & Enterprise respectively) state:

> "[The] potential impact [of big data] is so significant that it could transform every business sector and every scientific discipline. [...] The challenge of meeting the demand for skilled people, from both industry and academia, is one that is globally recognised. It is a challenge that cannot be tackled by government in isolation, which is why we will work with industry and academia to come up with solutions."
> (HM Government, 2013)

This discussion demonstrates the potential importance of analytics courses (alongside other training activities), to meeting the perceived skills gap associated with big data and analytics. Less clear, however, is the content that such courses should contain and their overlaps with other existing disciplines. A review of the current literature regarding analytics degree curricula suggests three major themes. Firstly, there are a variety of papers, many of which have already been discussed in this chapter, which seek to identify the disciplines that inform and overlap with analytics. Secondly, there are specific examples given of how a course can, or should, be constructed, and the topics and techniques that should be covered. Finally, there has been research analysing existing degree curricula, work that has the same intentions as this paper. Each of these last two will be discussed in sequence.

1.5.1 The Creation of Analytics Programs

Creating degree programs to meet the needs of analytics employers is a challenge that can be addressed in three ways:

1. Through modifying degrees in related disciplines to also incorporate some of the techniques, use cases and contexts of analytics. In other words, these would be traditional discipline degrees, albeit somewhat adapted to address some of the concerns of analytics employers and to incorporate some explanation of the role of that discipline in analytics practice.
2. Through creating specialisations within programs which offer some proportion of 'analytics' content, as well as that of a traditional discipline. There are several

examples of this approach, such as the University of Texas' MBA: Business Analytics Concentration.

3. Through creating bespoke "analytics" degrees, examples of which will be presented later in this paper.

In respect to the first of these, Chiang *et al.* (2012) discuss the potential for IS courses to evolve to meet this demand, and the extent to which this requires curricula to be developed, but also the focus to be changed. They identify a list of (primarily quantitative) modules that would need to be included, as well as a requirement to focus on "rapid interpretation and business decision making based on huge volumes of information" as opposed to an orientation towards "the management of transaction data and the production of information for management" (ibid., p 5).

The second approach, one which is closely related, is to provide specialisations or a collection of 'elective'-type modules, again within existing disciplines. An example of this is presented in Molluzzo *et al.* (2015), whereby the authors create a list of modules to include in such a concentration targeted at IS students. However, a caveat to such efforts, and those towards modifying existing degree courses, is given in Chen *et al.* (2012). Ultimately graduates of such schemes are more likely to still find roles within IS groups, albeit with a greater awareness of analytics and big data and its uses in organisations.

The alternative, of providing bespoke courses that combine aspects of the different disciplines involved in analytics, is less explicitly discussed in the literature. Partly this absence may be due to the political nature of such a task; ultimately this infers a 'shared ownership' between discipline groups. In other words, this is a trans-disciplinary task, and not necessarily one that suits the specialised nature of academic journals and research, therefore potentially limiting the opportunity to publish such work. However, it is a task seemingly important to understanding analytics courses, and to determining the proportions of each of the related disciplines that should be covered.

1.5.2 The Content of Analytics Curricula

The aforementioned study into the design of a 'data science' specialisation for IS students (Molluzzo *et al*, 2015) seeks to identify the modules that should be incorporated in such a program. Their research was based upon identifying recurrent topics over "the online syllabi of 21 introductory courses that contained Data Analytics or Data Science in their titles" (ibid., p 13). Notwithstanding the contribution of their work, there are two significant differences between their study and ours. Firstly, their area of focus is on data science specifically and on introductory courses; this research will seek to evaluate courses that have titles associated with terms such as “analytics”, “big data” and “data science” and based on full (graduate-level) degrees in the area rather than modules. Secondly our focus is on linking areas of analytics-orientated study to pre-existing disciplines (specifically OR); that is to better determine the root elements from which analytics degrees draw upon.

Another example of work in this area, is a series of surveys “to assess academia’s response to the growing market need for students with Business Intelligence (BI) and Business Analytics (BA) skill sets” conducted by the Association for Information Systems (Wixom *et al*, 2014, p 1). In this research, a large-scale questionnaire (n = 1,379) was conducted of university staff, students and practitioners. One area of exploration in the study surrounds the modules that educators currently offer at their university. Again, this presents some interesting results, however, such modules could be featured in a wide range of different degrees, and no indication is given as to whether these are electives or core units; what academic level they are at; or how they combine to offer the full range of required skills and understanding required of analytics graduates. Secondly, due to the survey instrument, there is a possible concern that different respondents may interpret the different categories differently, and therefore introduce some subjectivity into these findings.

1.5.3 Summary

In summary, the literature has demonstrated that this is a key area of study, and that the issue is problematised by the involvement of multiple disciplines in the analytics field, and the subsequent political issues and competition between fields this entails. However, to date, there appears to have been no significant empirical research into bespoke analytics courses and their contents, the traditions from which these modules are drawn, and, by implication, the relative importance of different disciplines on the analytics curricula. As such we posit the following, as the fourth and fifth objectives of this study:

RO4: To identify the academic disciplines with which analytics master’s degrees most closely align.

RO5: To identify the specific skills, subjects and techniques taught within analytics degree curricula.

These objectives are designed advance our understanding of the current academic landscape (in respect to the needs of analytics employers) at both a macro- (RO4) and a more micro-level (RO5). Whilst this, as indicated in this section, is an area of previous studies and discussions, it has also been evidenced that this remains an area of contention and one worthy of further investigation. Particularly, there is absence of any significant empirical research in this area, a gap the work will seek to meet.

Following from this, the first five objectives effectively tackle the individual facets of the overall problem. The first seeks to partition and define analytics in relation to other fields and disciplines; the second evaluates analytics research directions for the OR community; the third the requirements of analytics employers; and the fourth and fifth the current provisions for academic education in. However, whilst this can help assess the extent of the problem, a potential gap between current OR education and the needs of analytics employers, it does little to suggest how this may change. As the goal is to provide tangible recommendations that can be employed by OR and analytics educators, the final area of literature that will be analysed relates to both the potential barriers, and the examples of best practice that can be employed in curricula development. This will be the subject of the next section of the chapter.

1.6 Analytics and OR: Academic course design

The final area of inquiry, as detailed, concerns the development of curricula and some of the challenges this may present. The section begins by analysing curricula design in OR, and secondly other related disciplines, both through the form of case studies. Thirdly, we review some of the literature related to best practice in curricula development, and the challenges this may present for universities. These topics are discussed in sequence.

1.6.1 Curricula and Pedagogical Development in OR Courses

The previous section discussed some of the recorded efforts to develop bespoke analytics courses and modules. However, as noted, such developments can indeed be traced back further, through some of the disciplines that inform it. One of which, of course, is OR.

OR's history as a taught subject extends back to the early 1960s, with Lancaster University developing the first MSc in the subject in 1964. Mostly taught at master's level, the discipline has grown to offer degrees in universities all over the world. Beyond this, many of the students exposed to its methods would have done so as part of more general business and management degrees (such as MBAs). The content of such courses and modules has, however, been subject to some debate. For instance, as far back as 1970 research has been conducted into what such curricula should contain (e.g. Shannon and Biles, 1970), where the authors rank OR techniques based on use in practice.

However, beyond just the individual techniques and algorithms, the actual pedagogical approaches adopted have also been the subject of some debate. Cochran (2009, p 162) observes that the teaching or OR effectively splits into two separate eras:

> "Until approximately 1990 introductory operations research courses generally featured a heavy focus on the mathematical underpinnings of solution algorithms [...] while application-oriented concepts such as model building and the interpretation and implication of results were only briefly considered [...] Sadly (but in retrospect not surprisingly), most of the author's classmates did not share his enthusiasm. They left these classes seeing operations research as a collection of arcane mathematical tools that require massive computing power when applied to real problems and could only be used by highly technical individuals to solve extremely large and complex problems. In short, these students learned only to be intimidated by operations research."

The limited impact of such approaches eventually led, according to Cochran, to the questioning of the value of OR course and, ultimately, in 1991 the American Association of Schools and Colleges of Business determining that such courses should no longer be mandatory on business programs (a decision that was later reversed in 2003). In short, as described in Grossman (2001) and Grossman *et al.* (2016), there can be a disconnect between the mathematical orientation of the teachers of OR courses, and the orientation of most business students towards management issues and strategic change.

Since this period many educators within the OR discipline has sought to focus on new methods and approaches – specifically towards more situational uses, and applications of OR, rather than simply on the underlying algorithms and solvers (Cochran, 2009). Outside of the US, these ideas have been developed further, by a variety of authors advocating new approaches to the teaching of OR. Examples of this include the importance of real-world projects with real clients (e.g. Rand and Ranyard, 2013); encouraging more "inter-dependence" between OR teacher and student and greater student autonomy in their learning (Belton and Scott, 1998, p899); and the importance of case study materials (Bell and von Lanzenauer, 2000). Such methods emerged at a similar time to a movement towards problem-based learning across academic education as a whole (e.g. Birch, 1986).

Two examples of this new approach are given in Liberatore and Nydick (1999) and Robinson *et al.* (2003). Both relate to the redevelopment of OR modules in MBA programs (where arguably such problems are magnified due to the diversity and lack of mathematical background of some students) in the 1990s, with the former focused on a US university (Villanova University) and the latter in the UK (the University of Warwick). In both cases the authors detail changes that include:

- An empowerment of students to solve realistic problems rather than simply learning and applying specific OR approaches to arbitrary datasets;
- Integrated modules rather than "a survey that delivers a *technique of the week*" (Liberatore and Nydick, 1999, p 100);
- Discussing the context and process of approaches, and applications of techniques in real-world business settings;
- Encouraging the development of softer skills such as teamwork, communication and critical thinking.

In both cases the authors report these new approaches as successful, indeed Robinson *et al.* (2003) subtitle their paper "A Turnaround Story". However, that is not to say such changes have been universally adopted, nor that OR education is without problems. The aforementioned analysis of Sodhi and Tang (2008) identifies multiple problems in OR education, whilst Birge (2006) and Grossman *et al.* (2016) were still making calls for further moves towards new "active/co-operative learning" and "student-centred approaches".

With the potential challenge of new courses in "Analytics" or "Data Science", there is seemingly enough scope for OR educators to at least seek to evaluate the current state of their field, and the potential requirements for updates to curricula and pedagogical methods. However, they are not the only discipline that may seek to do this, and there is obviously the potential to learn from the activities of others. As such, the next section will evaluate instances of change and evaluation in other related disciplines.

1.6.2 Curricula and Pedagogical Development in Related Disciplines

To add to the previous example, this section will look at two alternative disciplines, and a brief case study of developments in statistics since the 1990s, and the development of "a set of recommendations for four-year programs in Information Technology" (Lunt *et al*, 2008, p 12). Such examples not only provide a comparison point for the OR discipline, but also relate to disciplines that have strong linkage to analytics and its curricula.

1.6.2.1 Statistics and the Drive to Data

Again, in the mid-1990s, as a part of a wider movement towards the "democratisation of mathematics" (Vere-Jones, 1995, p 13), many academics and educators began to question the effectiveness of the content and pedagogical approaches of university statistics courses. As observed in Matthews and Clark (2003), even many successful students of statistics courses were unable to explain the underlying 'meaning' of common statistical metrics (such as the mean), but had merely memorised the set of steps involved. A rather damning indictment of the state of affairs was given in Gould (2010, p 298):

> "By the 1980's it was clear to a growing number of statisticians and educators that attempts to teach statistics to a general audience were failing miserably [...] The curriculum had become focused on teaching procedures and rote memory. Homework problems were dull, tedious, and, by using idealized contexts, failed to teach students the usefulness or applicability of statistics to real world problems."

In part, the proposed solutions were not dissimilar to those of the above case study of section 1.6.1 (e.g. Moore 1997, p 127) of fostering active learning and a greater emphasis on problem-solving over memorised algorithms. However, equally important according to many was a need to emphasise data over techniques; as argued in Gould (2010, p 298), summarising the perspective of Denning, "statistics might *use* math, but it was *about* data".

Quite how to do make this emphasis though was a matter of some debate. Fundamentally, and to paraphrase George Orwell, all statistics uses data, but (the argument goes) some data are more useful than others. Stedman (1993), Stork (2003) and Carnell (2008) advocate the use of student-generated data such that the source, purposes and contexts of the input are more clearly understood. Gould (2010) argues for the use of "real" data; not just to mean non-artificial datasets, but to mean datasets that speak to the student's experiences, that are 'real' to their lives. Lesser (2007) argues for data relating to social justice, potentially empowering students to use statistical learnings to seek to impact real world change.

In each case the authors find successes in these new methods, with increased student engagement and improved learning outcomes. However, that is not to say that the battle is now won. As argued in Hardin *et al.* (2015), the advent of analytics and data science means a 'datafication' of statistics syllabi is even more important, and along with the addition of more computing skills, degrees in the subject need to become even more data focused, particularly in recognising the "3 V's" of big data (as discussed in section 1.2.1), volume, velocity and variety, aside more traditional datasets and problems.

1.6.2.2 Designing the IT Curriula

The second case study centres on the creation of curricula guidelines for undergraduate degrees in IT, undertaken by the Special-Interest Group for Information Technology Education of the Association for Computing Machinery (detailed in Lunt *et al*, 2008, p 12). The underlying rationale for this endeavour was that university teaching of computing had grown to such as size that it needed to be considered as a collection of sub-disciplines (determined by a joint task force of members from the ACM, AIS and IEEE-CS to comprise of Computer Engineering, Computer Science, Information Systems, Information Technology and Software Engineering). Accordingly, each of these five categories was the subject of further analysis and, ultimately, curricula guideline report was created for each. For this particular case study, somewhat arbitrarily, IT was selected (Shakelford, 2005).

The report on IT curricula was some 5 years in the making, and included focus groups of more than thirty participants (Lunt *et al*, 2008). The principal outputs of the report included: a body of knowledge (including 85 items); learning outcomes for each of these; a subset of 81 or these 85 which are considered core; a list of advanced outcomes beyond the cores; curriculum models (suggested approaches to covering the elements of the curricula); and associated course descriptions (ibid. p 6).

Whilst the outputs, as this would indicate, are far too wide reaching to report in full in this thesis (the report totals 139 pages), there are some key aspects. Most importantly (for our purposes) this is at the most general levels. The authors determined that the discipline has emerged as a response to demand for graduates to work in the industry, and therefore a logical starting point is the requirements of employers (as has been recognised in this research methodology). From this 14 key skills and aptitudes were identified (Lunt *et al*, 2008, pp 18-19). Further to this, these skills were consolidated, reaching the representation of the core curricula elements shown as figure 9.

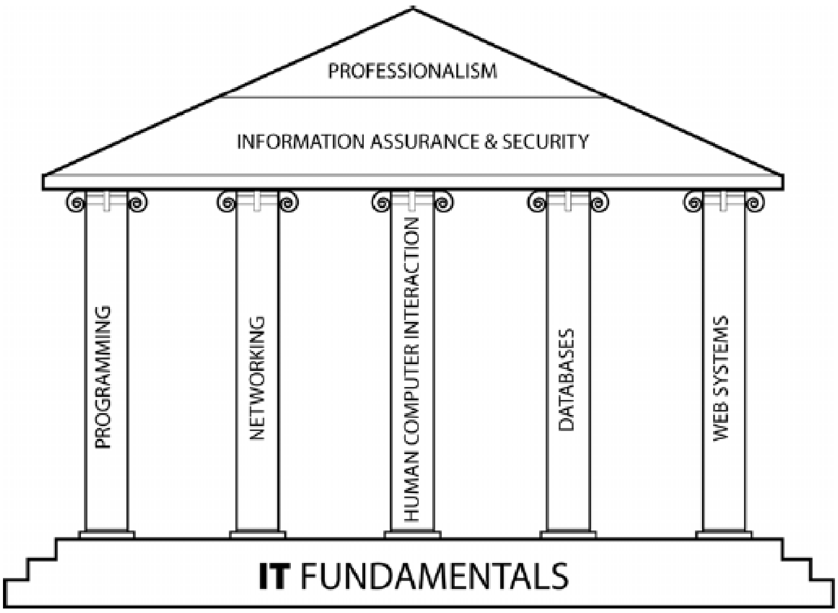


Figure 9 The Information Technology discipline (Source: Lunt *et al*, 2008, p 19)

The final product produced in this study (and the other four related reports on the other sub-disciplines) is a comprehensive curriculum guide for which undergraduate programs can follow. There are of course clear benefits to this in aiding course-designers and ensuring a consistent skillset across the graduates such course develop. However, there are certain caveats.

Firstly, and most obviously, this represents a very vocational approach to curricula design. As such this taps into a fundamental question about the purpose of academic courses; do they represent a factory-line designed to produce 'out-of-the-box' employees for specific vocations, or are they there to introduce an academic discipline, and its theoretical and less practical aspects. In reality, the answer is typically somewhere in the middle and somewhat dependent on the nature of the discipline – one would assume a course in Information Technology to be far more practically-orientated than a course in Philosophy. As *a priori* assumption, it may be expected that master's level course in OR and analytics are likely to be more towards the vocational-end of the spectrum, but possibly not to the same extent as with this case study.

Secondly, such an approach is dependent on a reasonably comprehensive understanding and consensus on the core skillsets are for analytics graduates, something this chapter has shown not to be the case. As such, obtaining the level of granularity presented in this report may be an unrealistic aim; or indeed, if there are multiple directions such courses can take, may be overly restrictive. However, some guiding principles and core components, as are shown in figure 9, would be a sensible target.

1.6.3 Challenges and Best Practice in Curricula Design

The final part of this section, and of the literature review as whole, concerns prior work and guidelines as to how curricula and courses can be designed. Whilst there are examples of this in the literature, there is perhaps less than may be expected. This is, however, perhaps not completely unsurprising. As identified in a survey of US academics, often teaching can come second-place to research activities in academia, with one respondent arguing: "although considerable lip service is paid to the importance of teaching [...] the more attention one pays to the real needs of the students we teach, the more the lip service and the fewer the rewards" (Gray *et al*, 1996, p 65). Although perhaps not really within the actual scope of this research, such issues are worthy of note as they have clear implications for the actualisation of master's courses in analytics and the updating of OR curricula. Inasmuch as the research seeks to create 'best practice' for such courses, it is important to also recognise the limitations and potential barriers that may impact their implementation.

Some such issues are identified in one of the better-known texts in this area, *Designing and Assessing Courses and Curricula* (Diamond, 2008), from which figure 10 is drawn. A principal element in this model for program development is what is labelled "Project-Specific Factors". Within this there are administrative concerns, such as accreditations and student-specific restrictions. However, there are also resource issues, in both financial and human terms. As alluded to earlier in the chapter, this latter issue can affect this area in two distinct ways. Firstly, that analytics seemingly require a wide-ranging skillset of graduates, skills which need to be matched by the knowledge available teaching resources. Secondly, and effectively the reverse of this, for an institution seeking to develop analytics programs (or re-develop OR programs), the shape of these are likely to be influenced by the modules already run at the university. This is manifest both in the modules that will be offered to students, particularly in respect to elective modules, but also may represent 'no-go zones' if other degrees and faculties or schools of the university already have similar provisions.

On the best practice-side of the debate, Diamond's model emphasises the importance of recognising the multiple facets of course requirements: the needs/desires of students; of society; the current directions of research; educational priorities; and that of the discipline as whole. This re-enforces that the output of any recommendations too need to be multi-faceted, and to recognise that the destinations, and future contributions, of graduates will not all be the same.

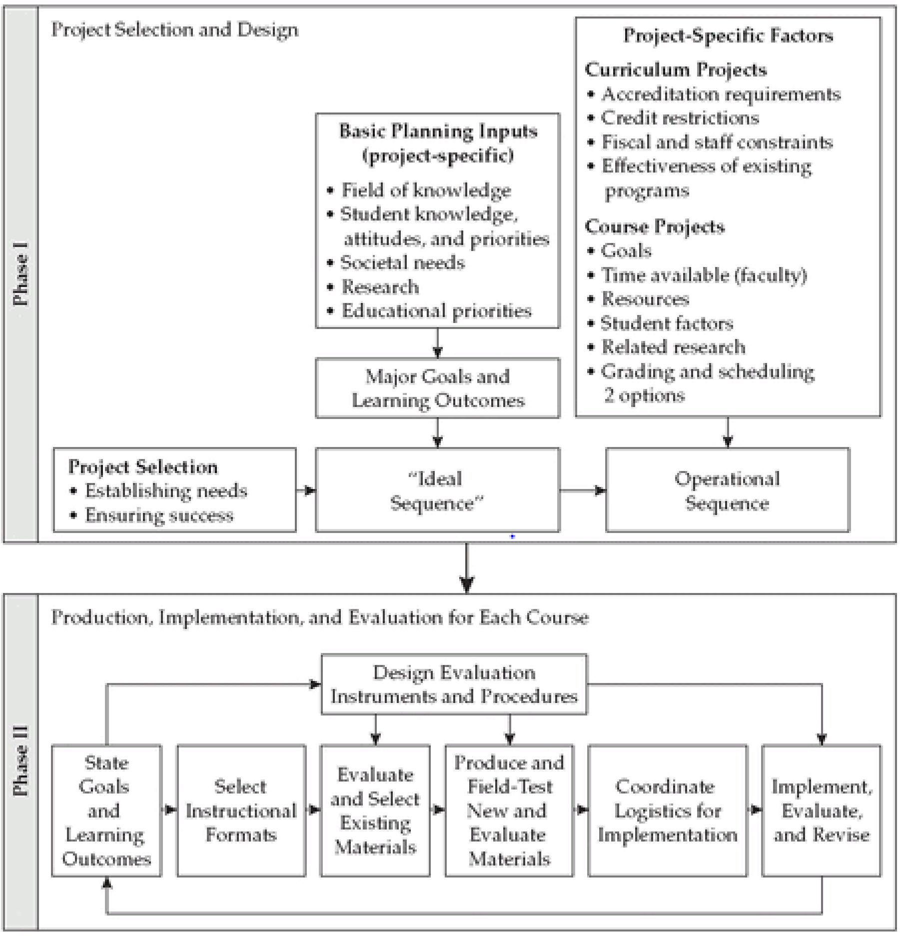


Figure 10 Process for the development of educational programs

(Source: Diamond, 2008, p 1226)

1.6.4 Summary

In summary, this section has demonstrated that the central purpose of this research, effectively identifying best practice for analytics and OR education, needs to go beyond considering just one stakeholder (the potential employers of such graduates) and over one aspect (course curricula). Whilst these, of course, are critical concern, also of issue is the pedagogical aspects of teaching such degrees, the organisational pressures and concerns of universities as wider institutions, and the need to engage students in the materials.

On this basis, we posit a sixth research objective, which in combination with the previous objectives, allow us to reach our seventh and final objective:

RO6: To identify the potential barriers and concerns that impact the creation of analytics and OR curricula

RO7: To create a framework for the development of analytics and OR degrees.

The remainder of this chapter will provide a summary of its contents. Thereafter, the research objectives are recapped, and the structure of the remainder of the study is detailed.

1.7 Summary, Objectives and Research Structure

Over the course of the chapter we have presented both an introduction to the research, as well as a literature review of the key areas of previous study. These incorporated the growing interest in analytics; the directions and shape of research into analytics and the lack of OR research in this area; studies into the skills requirements of analytics professionals; the development of analytics degree programs; and the issues and concerns of developing academic curricula. In doing so the following research objectives have been formulated:

RO1: To determine the relationship between academic definitions of analytics, operational research, and other related fields and disciplines.

RO2: To develop a research agenda for the OR community which addresses the concerns associated with analytics.

RO3: To determine the skills requirements of analytics roles and the extent to which these may be met by OR professionals.

RO4: To identify the academic disciplines with which analytics master's degrees most closely align.

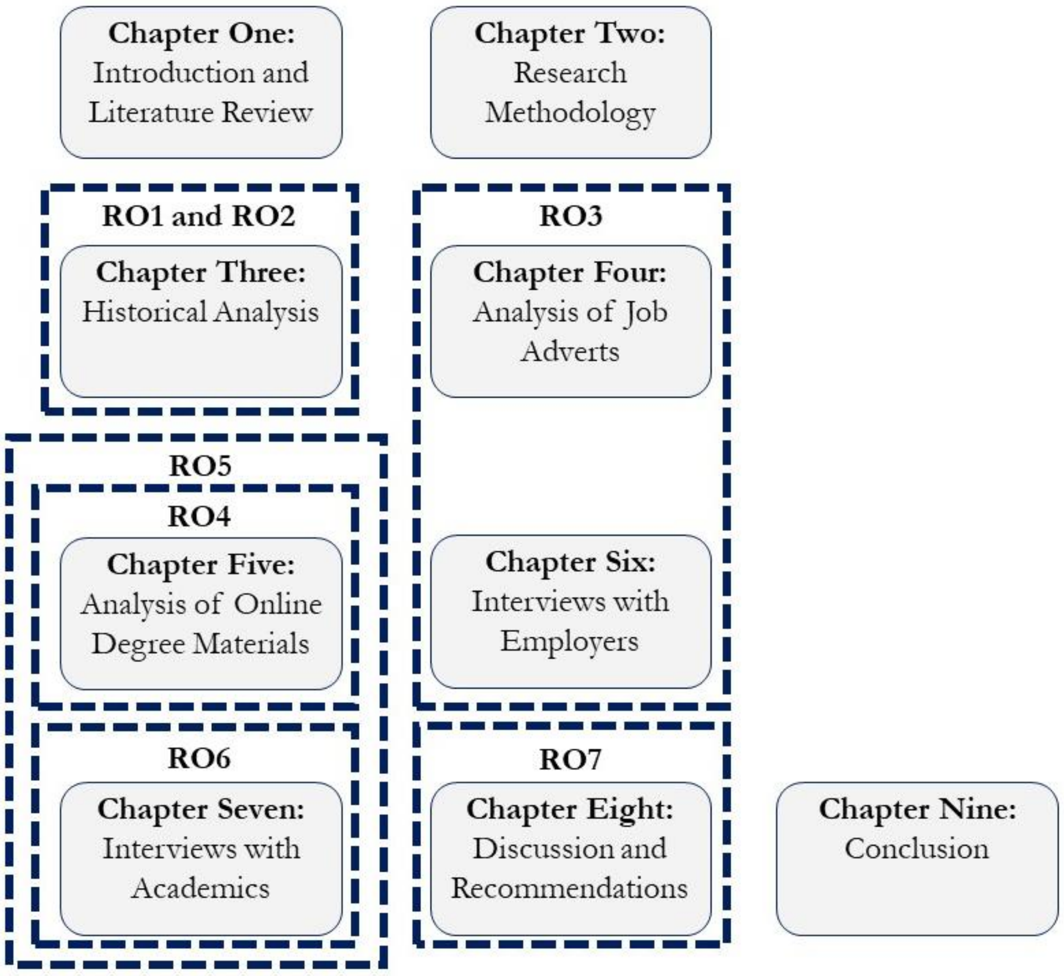
RO5: To identify the specific skills, subjects and techniques taught within analytics degree curricula.

RO6: To identify the potential barriers and concerns that impact the creation of analytics and OR curricula.

RO7: To create a framework for the development of analytics and OR degrees.

These objectives form the structure of the remainder of the thesis, with a variety of research methods employed to address them. The second chapter will detail these methods and the overall research approach. The third chapter employs a historical analysis designed to address RO1 and RO2. Chapter four presents a topic model analysis of analytics job adverts and, in part, address RO3. The fifth chapter employs further text analytics approaches to evaluate online degree materials associated with analytics, OR and related degree courses (addressing RO4 and partially addressing RO5). Chapter six presents the results of a series of interviews with analytics and OR employers, to complement the analysis of chapter four to address RO3. Chapter seven presents analysis of further interviews with academics and the developers of analytics degrees to fully addressing RO5 and RO6. The eighth chapter seeks to synthesise these varied results to create recommendations and a framework for the development of OR and analytics degrees, addressing RO7 and the core goal of the research. Finally, chapter nine provides a conclusion to the work, a summary of its contributions, and discussion of the limitations of the study and opportunities for future research. This structure is summarised in figure 11.

Figure 11 Structure of the thesis and research objectives



As indicated, there are a wide range of research instruments employed to meet the goals of this research. Accordingly, the next chapter will detail these further, and explain how they fit within the philosophy of the research.

2 RESEARCH METHODOLOGY

The previous chapter has detailed the purpose and positioning of the research, introduced much of the prior literature in relevant areas, as well as the research objectives and thesis structure. Building upon this, the next concern is presenting the research philosophy, methodology and instruments used. In accordance with our positioning of the research, as a multi-faceted problem with multiple considerations, the research methodology too has multiple facets. These combine not only the concerns of the multiple stakeholders involved in this space, but also multiple research paradigms (in the tradition of mixed methods research – MMR).

In doing so we recognise value and issues in both positivism (the belief in research that is, and can be, scientifically verified) *and* interpretivism (whereby the argument is that in the socially-constructed world, such objectivity is generally unattainable and overtly simplistic). More specifically, as detailed over the course of this chapter, we seek both some of the methods of scientific testing afforded by positivism, as well as recognising that this research area is from within the social realm, meaning an appreciation of the traditions and methods of interpretivism may be key to fully developing an appropriate response. In other words, we posit that such flexibility is a necessity in managing such a wide-ranging problem space, and therefore key to reaching the goals of this work.

On this basis, the chapter is arranged as follows. The first section further discusses the research philosophy. Section two details the research strategy and approach. Sections three to seven detail the four main strands of research (namely a historical analysis; a text analysis of job adverts; a text analysis of online degree materials; and two sets of interviews), detailing overall approach and analysis strategies. Finally, section seven presents a brief summary of the chapter.

2.1 Research Philosophy

As detailed in the introduction, the approach of this study is of mixed methods research (MMR). The natural philosophical conjugate of such an approach is a pragmatic stance, and this indeed represents the position that is taken. Whilst pragmatism is not necessarily new, indeed it is typically traced back to the 19th Century and exponents such as Charles Sanders Peirce, William James and John Dewey (Hookway, 2016), it is within the last thirty years that the approach has returned to prominence in academic research (e.g. Johnson and Onwuegbuzie, 2004).

Pragmatism represents an inherently practical approach to scientific enquiry, whereby considerations of outcomes are typically championed above theoretical dimensions, in contrast to the theory generation and theory testing methods associated with interpretivism and positivism. In other words, whilst we seek theoretical contributions, the practical contributions this work may produce, are given greater weight. That is to say we choose to more greatly emphasise the benefits to educators and employers our solution may bring, than any contribution to the wider academic knowledge.

Research pragmatism as a philosophy has obvious benefits for such research designs, allowing a flexibility to employ the different methods to unpick multi-faceted problems (such as has been described here). However, it is a philosophy that has received criticisms. Cameron (2011, pp. 97-98) argues these can be considered in five categories:

- **Paradigms:** that pragmatism is effectively without a paradigm (eclecticism), and, for 'purists' at either end of the paradigmatic debates, therefore lacking as a fully functional research philosophy;
- **Pragmatism:** in a related way, a potential concern is of a "short-sighted practicalism" (ibid., p 97) whereby the research is overly concerned with its practical implementation at the expense of its theoretical underpinnings;
- **Praxis:** relating to the issues that affect the integration of research instruments, data and analysis methods, often complicated by their variety in MMR;
- **Proficiency:** another implication for the researcher is an increased requirement for competency in two, often very different, skillsets (quantitative and qualitative methods);
- **Publishing:** a final concern, albeit one that is of less concern to this thesis, is finding the appropriate publishing venues for such research.

In effect, ignoring the fifth category of publishing, these criticisms can, to some extent, be collapsed into two: philosophical concerns ('paradigms' and 'pragmatism') and practical concerns ('praxis' and 'proficiency'). In the main, the practical concerns are addressed later in the chapter as the research methods are detailed in full. However, it is worthwhile considering the philosophical concerns as they relate to this research.

Although 'practicalism' is deemed to be a criticism of such an approach, practicalism may also have a philosophical underpinning in and of itself. Ultimately problem-types do need to be considered in the choice of research methods, and therefore should have an implication on the philosophy employed. Indeed, for the purists of either of the traditional research paradigms, a 'paradigm-first' approach can lead to the trap described by Abraham Maslow (1969): "I suppose it is tempting, if the only tool you have is a hammer, to treat everything as if it were a nail". In the instance of this research, as alluded to in earlier discussion, the specific nature of the problem does suggest a need for a MMR approach. Firstly, this area is one of considerable uncertainty (for instance, the confusion about OR's relationship with analytics as demonstrated in figure 2), suggesting a need to employ interpretivist methods to understand the subjective opinions of different stakeholders. Secondly, however, there is a need to try to find some degree of uniformity and recurrent patterns across the breadth of employer requirements and university provisions, suggesting the value of quantitative research.

Such a position could be described as "situationalist" (Cameron, 2011, p 100), whereby the situation dictates for the researcher which method is used. Whilst we may argue that such a position is perfectly valid in this case, it does still open the research up to potential criticism as to its lack of clear theoretical underpinnings. In response to this, the research is positioned beyond the situationalist position, to one where the pragmatist philosophy is considered more than a mere convenience or a problem-specific choice. In effect, we invoke the "dialectic stance" presented in Teddlie and Tashakkori (2010, p 15), described as "[assuming that] all paradigms have something to offer and that use of multiple paradigms in a single study contributes to greater understanding of the phenomenon under investigation". In effect, this translates as saying that there are philosophical merits in both quantitative and qualitative merits, and rejecting that the embracement of one necessitates the exclusion of the other.

It could be argued that this cannot be the case, as some element of each tradition is incompatible with the other. However, to counter this we may also argue that these elements are inaccurate or inadequate; to effectively invoke the opposite of the dialectic stance, perhaps best demonstrated by the famous assertion that "all models are wrong, but some are useful" (Box and Draper, 1987, p 424). In other words, we posit that both paradigms have value and add to the researcher's toolkit, and any argument of incompatibility comes from the inadequacies of both paradigms.

Despite the plutocratic nature of MMR, as many observers note, inasmuch as researchers will typically have a predominance of skills in one of the traditions of qualitative and quantitative research, much research of this kind too will have a stronger flavour of one over the other. Whilst, in the spirit of full disclosure, the author would consider himself to be stronger in the quantitative area than the qualitative, that is not to say that this research particularly values one approach. Indeed, to some extent doing so could be argued to fall somewhat into the traps of the paradigm wars ("quants with a side order of qual").

A more useful conception, certainly in the contexts of this research, is not to label the 'flavour' of MMR based on the positioning of the methods, but moreover on the positioning of the pragmatic philosophy. In Johnson *et al.* (2007, p 125), three types of pragmatism are presented, that of the right (where a realist philosophy is dominant to a pluralist philosophy); of the left (where the opposite is true); and of the centre (where the two are balanced). Using this language, a 'centre-right' position would best describe the philosophy taken in this work (and indeed would best summarise the overall position of the author). That is to say that whilst an acknowledgement is made of the lack of true objectivity in much of the social realm, that does not necessitate that realist methods are automatically 'wrong' or ineffective. Indeed, there is considerable evidence that positivist-type methods can bring tangible benefits, even in socially and politically charged problem-spaces. Whilst interpretivist-type methods can be also utilised in such a 'centre-right' stance, the adoption of such a position, as envisioned here, is to seek a more realist methodology except where this is detrimental to do so. To borrow again from the imagery of Maslow, the preferred tool is the hammer, but we endeavour to verify that the problems we are aiming it at, are in fact nails.

In other words, whilst by default the MMR position "rejects traditional dualisms" (Johnson and Onwuegbuzie, 2004, p 18), particularly in respect to a divide between realism/antirealism and objectivity/subjectivity, the adoption of a (quasi-)objective framework has clear benefits in terms of reaching relatively firm recommendations. That is not to say that the subjectivity of the area is not recognised, nor that there is a genuine belief that the research will capture some unequivocal "truth", but moreover a recognition that there are practical benefits in aping something of the positivist tradition. This is exemplified in our approach to coding interviews where *a priori* codes, in part drawn from quantitative research, are used to initially begin this analysis (further detailed in section 2.6).

However, the adoption of any research philosophy has implications and potential limitations, and the pragmatic approach employed in this study is no exception. To demonstrate how these are managed, the next section details the research strategy employed.

2.2 Research Strategy

The specific research strategy can be illustrated by positioning it on Johnson and Onwuegbuzie's mixed-method research design matrix, shown in figure 12. In this design, the authors use the capitalisation of the terms "QUAN" and "QUAL" to indicate the 'paradigm emphasis decision' – that a capitalised method is the principal instrument, whereas lower-case listings of the terms indicate secondary or supplementary instruments. To distinguish between the different "time order decision[s]", Johnson and Onwuegbuzie use "→" to denote one instrument as building on the findings of another (in comparison to "+" denoting concurrent research methods).

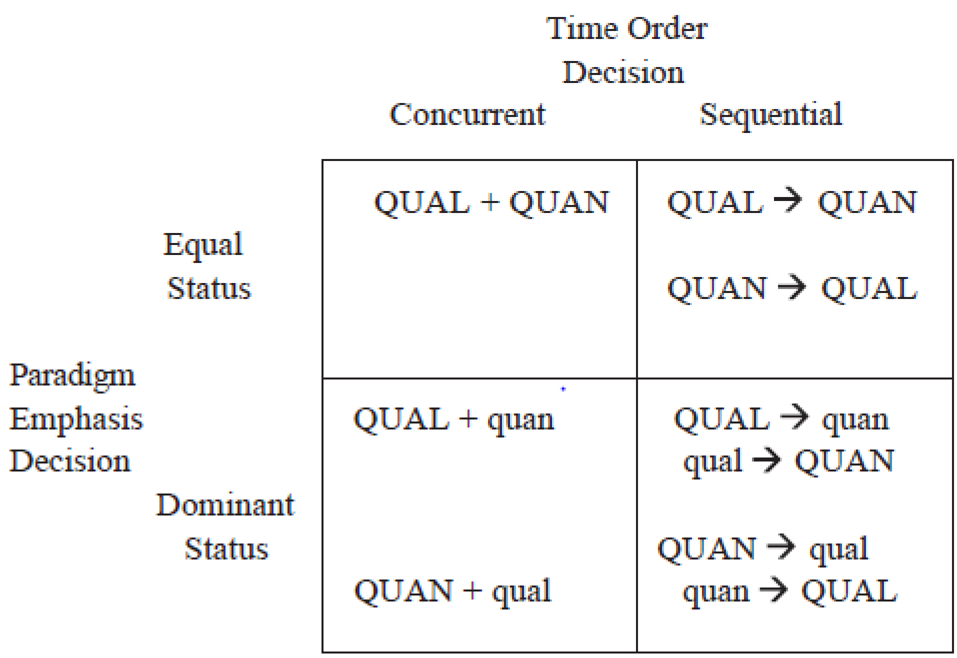


Figure 12 The mixed-method design matrix

(Source: Johnson and Onwuegbuzie, 2004, p 18)

As stated previously, despite adopting a 'centre-right' approach to the pragmatist philosophy, the use of qualitative and quantitative are given equal weight in this work. Therefore, to use the notation of figure 12, the research described includes three "QUAL" instruments, and two "QUAN" instruments (intentionally capitalised to show no method to be subservient to the other). However, in respect to "time order decision", the research mixes both sequential ("→") and concurrent ("+") methods. In other words, and whilst the methods are sequential in the work, there are two methods in the design which are used independently of the other.

To be specific, the research progresses as follows. Firstly, a qualitative analysis is performed of secondary data. In part, an output of this is a better understanding of analytics and its relationship with other disciplines. This understanding then forms an input for the second research method, a quantitative analysis of analytics job adverts in relation to those of other disciplines, and the third, a parallel (quantitative) comparison between the online materials concerning analytics degrees and their curricula, and those of OR and related disciplines. These second and third methods are effectively concurrent (although presented consecutively). The results of the second method (the job advert analysis) is used to inform the fourth (interviews with analytics/OR employers) and these findings, along with the results of the third method (analysis of degree materials), inform the fifth and final method. The progression is shown in figure 13.

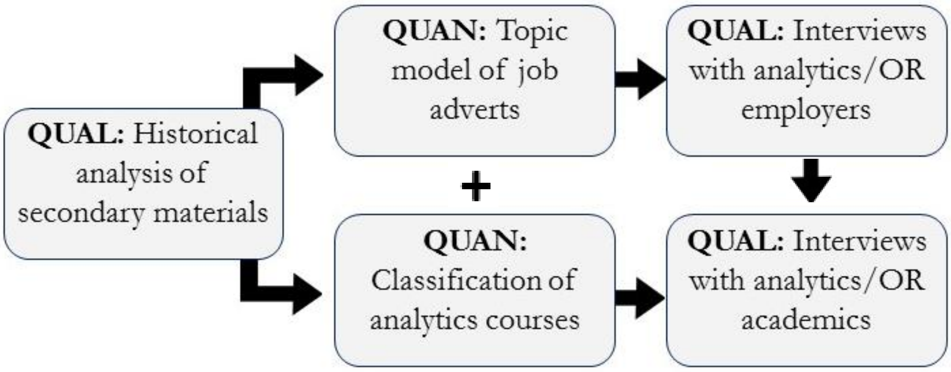


Figure 13 The use of quantitative and qualitative instruments in the research design

It is important to note though, that despite the indications of a sequence to the research ("→"), in many ways the approach describes differs somewhat from the model of Johnson and Onwuegbuzie. In effect, each work package is relatively 'stand-alone' and not specifically designed to supplement each other but to discover independent insights. Whilst the thesis brings these together, and in their combination, finds further insights and more complete answers to the research question, equally each can more-or-less be regarded as discrete pieces of work (with the possible exception of the fourth and fifth instruments; the interviews). Such a design has benefits, as it allows for methods to be matched to facets of the problem, but also presents risks in respect to consistency and 'theoretical completeness'. However, again this really is just a trade-off between practical contribution (on the basis that more targeted work packages may better capture the specifics of each subset problem) and theoretical contribution (in that the different elements are more complementary than they are conjoined). In other words, we consider that this decision does remain consistent with the stated philosophical positioning of the research.

Moreover, again we draw from this parallels with two of the key subjects of this research, analytics and big data. As detailed in section 1.2.1, one of the key defining traits of big data is the ability to do things, or ask questions that are not possible in traditional scientific analysis. In the same spirit, the slightly non-traditional approach utilised in this research, some of which utilises 'big data' sources (i.e. text data), is hoped to be able to derive insights that may not have been available in a more typical methodology. In many ways, we seek to emulate the idea of using multiple weak-indicators that in combination out-perform (predictively) single, more robustly gathered data sources (e.g. Anderson, 2008). The research instruments used are more singular, and less cohesively conjoined as they may be in a more traditionally methodology, but the goal is that collectively they may outperform similar research using just a single method.

2.2.1 Validation and Limitations

Although, as described, many of the research methods are effectively 'stand-alone', through the process of triangulation, the subsequent research instruments (figure 13) also act as validation methods for their predecessors. This aspect has an important contribution to the research, particularly in consideration of the ambiguities of the area and the complexity of the problem definition. However, despite this validation, it is important to note that the research does still have limitations and assumptions (implicit and explicit).

Firstly, as discussed in section 1.2, there are significant concerns regarding the nomenclature of the area. Although an understanding of this develops over the course of the study (particularly in chapter three), some assumptions/simplifications have been made in this regard. In terms of the OR discipline, numerous debates have occurred as to the relationships and classification of the broader discipline, particularly around terms such as "operational research", "operations research", "management science" and "decision science", as well as a perceived division between "hard" and "soft" methods. In the interests of clarity, and to avoid protracted debates about precisely which elements are associated with which of these terms, the term "OR" is used as the superset of all of these, and broadly these are treated as synonyms in this context.

A similar scenario occurs with the range of terms related to analytics. Despite the relative infancy of this area a wide variety of terms are in common use, including "analytics", "business analytics", "decision analytics", "data analytics", "big data analytics", "data science", "big data science" and others. Although understanding these terms and their potential differences is, in effect, one of the aims of the research, *a priori* no distinction is made between them, with again "analytics" representing the *de facto* superset for each.

Secondly, there is a concern regarding the breadth of the scope. Whilst universities all over the world are developing courses in this area, our research question specifically focuses on UK universities. However, inasmuch as this restriction has some logical sense, in the modern age of globalisation, it is naïve to think that the development of courses outside the UK has no effect upon the courses within it (e.g. Blight *et al*, 2000). This is likely manifest in the movement of academics internationally between institutions, an increasingly internationally-orientated student base, and a recognition of the variety of destinations of graduates of such courses around the globe. Whilst these concerns are, to some extent, considered beyond the scope of the research, there impact needs to be recognised as a limitation of the research.

An additional design choice in this regard, is to limit investigations to master's-level courses. This effectively ignores the growth of bachelor's-level courses in such areas such as Lancaster University's *BSc in Business Analytics & Consultancy*, and the University of Warwick's *BSc in Data*

Science. Of course, such programs are entirely consistent with the 'spirit' of the research, and may also impact on curricula design where students embark on a pathway between undergraduate and postgraduate study. However, in the interest of creating a narrow enough scope, and ensuring some consistency can be given in the recommendations of the study, they have been excluded in the main from consideration, a simplification which is too a limitation. Finally, of course, each of the individual research instruments is subject to its own limitations and assumptions, which are discussed in the relevant sections of this chapter.

2.2.2 Summary of the Research Strategy

In summary, this section has described a mixed method strategy with multiple work packages designed to meet different research objectives. A summary of this is shown in table 12, before the remainder of this chapter will describe the details of each.

Table 12 Association of work packages to thesis contents and research objectives

Thesis chapter	Methodology section	Work package	Research objectives
Three	Section 2.3	Historical analysis.	RO1; RO2
Four	Section 2.4	Topic model analysis of "analytics" job adverts.	RO3; RO5
Five	Section 2.5	Support vector classifier and module analysis of "analytics" degree materials.	RO4; RO6
Six	Section 2.6	Template and matrix of interviews with "analytics" employers.	RO3; RO5
Seven	Section 2.7	Template analysis of interviews with "analytics" course designers and educators.	RO4; RO6

2.3 Historical Analysis

As part of the CLR analysis of the existing literature, two shortcomings were identified. Firstly, a lack of clarity as to the relationship between analytics and OR, as well as with other related terms such as business intelligence (BI), decision support systems (DSS) and others (a shortcoming we align with RO1). Secondly, that there is a paradoxically little research into analytics from the OR community (aligned with RO2). The historical analysis which forms chapter three is designed to meet these shortcomings and objectives, the methodology of which is detailed here.

The underlying rationale for the approach is that one logical reason for the similarities between analytics, BI, OR, and some of the fields discussed is that fundamentally they all share a similar purpose: the improvement of business operations and decision making through the utilisation of information, quantitative analyses, and/or technologies. However, rather than mere coincidence, an alternative interpretation would be that they are all components of a larger, and broader movement, which, we argue, has had significant effect on the practices of management for some considerable time. This movement, using the concepts introduced by Kuhn (1962), can therefore be described as the dominant paradigm in the 'science' of business management.

Though it has precursors, particularly Adam Smith's *The Wealth of Nations*, the second industrial revolution (c1867-1914) can be seen as the main catalyst for the inception of the paradigm, a "paradigmatic shift" in Kuhn's terminology. In the new industrialised cities of the early 20th century the ideologies of scientific management, mostly attributable to the work of Frederick Taylor, came to prominence. The approach championed the use of statistics, efficiency, rationality, and the application of science to the problems of process and people management. Whilst the movement's momentum eventually waned, it had significant impact at the time, as well as leaving a clear legacy on management practice (Taksa, 1992). Accordingly, it would seem appropriate to consider this new approach as the start of a new management paradigm. Not only is there the notion of "inconsummerability" with the practices of proceeding periods, but also that there has been the progression of "normal science" in the years since (Kuhn, 1962).

This is supported by the work of Locke (1989) into what he regards as the start of a new academic paradigm at a similar time. He argues this brought a new approach of management training through education, opposing the tradition of coming up the ranks from "apprentice" to "master-craftsmen", a practice he argues as being without "applied science" (Locke, 1989, p 4). The argument here is that the stimulus for this paradigmatic shift in management training is preceded by a paradigmatic shift in attitudes to the practice of management; the latter of which being the focus of this analysis.

The proposed management paradigm will be labelled *dianoetic management*: dianoetic being defined in the Collins English Dictionary as "of or relating to thought, [especially] to discursive reasoning rather than intuition". The term, although somewhat obscure, has the benefit that it does not have the connotations with pre-existing terminology (e.g. scientific- or analytical management). However, the meaning is appropriate to the practices and purposes of the paradigm: the development of management based upon logic and evidence rather than 'gut-feeling'. This is not to reduce the importance of intuition, which still has an integral and essential role in effective decision making. The advances and applications of the paradigm have sought to make available data, tools and analyses to provide the evidence to allow decision makers access to discursive evidence that that can supplement their use of intuition and experience for more effective decision making (see Shah *et al.* (2012) for further discussion on this area).

This chapter will seek to analyse the dianoetic management paradigm through analysing its historical development. Having defined the object of the study and the timeframe included (from

1910 to the present day), a remaining concern is the sources to use. The paradigm clearly incorporates a wide range of traditional academic disciplines, as highlighted in the earlier discussions about those that inform analytics. These can be summarised as fitting into one or more of the following categories:

- **Technological:** incorporating the various tools used such as hardware, software, and networks, which together support the efficient processing of data.
- **Quantitative methods:** the applied quantitative approaches to analysing business data, such as statistics, machine learning, econometrics and OR.
- **Decision making:** the tools, theories, and practices used to support and understand the decision-making process. This inherently interdisciplinary area is incorporated into many academic traditions, most obviously in psychology and behavioural science, but also in many of the other disciplines of the paradigm (e.g. human-computer interaction and visualisation in information systems, or problem structuring methods in OR).

Based upon this categorisation a taxonomy has been created in figure 14, incorporating the disciplines each contains. Each includes disciplines that are effectively located in just one category, such as electrical engineering (technologies), mathematics (quantitative methods) and psychology (decision making). Contrastingly, some disciplines can be considered part of more than one category. Machine learning, a branch of artificial intelligence, has both technological and quantitative components. Information Systems, the study of the use of information technologies in organisations, has obvious connection to computing (technologies), as well as behavioural studies linked to decision making.

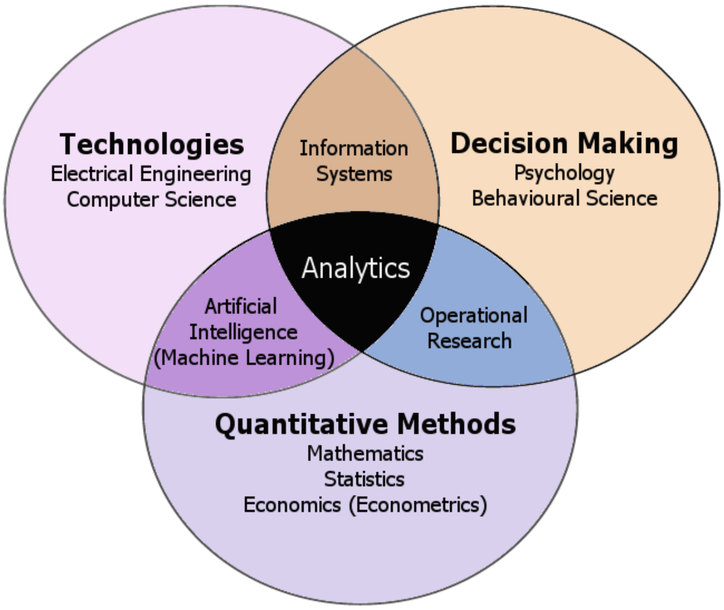


Figure 14 Dianoetic management in 2016: A taxonomy of disciplines related to analytics

Finally, OR, which has a clear quantitative aspect, has evolved to include focus on the more subjective areas of decision making. This is particularly evident in 'soft OR' (Rosenhead and Mingers, 2001) and 'behavioural OR' (Hämälläinen *et al*, 2013), but also, in approaches such as multi-criteria decision analysis (MCDA), the use of more subjective expert or decision maker judgement as a data input (see Köksalan *et al.* (2011) for further discussion of the development of these methods). Indeed, arguably it is this focus on decision making and decision makers that differentiates the discipline, in both its 'hard' and 'soft' variants, particularly its use in practice.

As these disciplines are argued to be a relevant part of analytics, and therefore the dianoetic management paradigm, they should be a relevant part of the recording of its history. Accordingly, sources from each of these academic and practitioner traditions will be evaluated alongside developments in data processing and management, as essentially each of these disciplines can be seen to be dependent on the consumption of data (albeit qualitative data in some cases) and each, at least in their use in business contexts, is typically used to support business management.

The final aspect of this analysis is the division of the history into periods. This serves two purposes. Firstly, it is an abstraction allowing the history to be 'shaped' into segments, and then more easily analysed. Whilst there is some arbitrariness to abstractions of this kind, the periods do demonstrate specific characteristics. Secondly the periods chosen reflect the years in which the different fields were particularly prominent. The paradigm will be divided into six periods:

1. **Scientific Management**: the years between 1910 (the publication of Taylor's monograph The Principles of Scientific Management) and the end of the Second World War.
2. **The Scientific Method**: the period between the end of the war and the mid-1960s, marked by the increased use of OR in businesses.
3. **Management Information Systems**: the mid-1960s to early-1970s, characterised by the growth of management information systems (MIS).
4. **Decision Support Systems**: the early-1970s to late 1980s when DSS were particularly prominent.
5. **Business Intelligence**: the early 1990s to the mid-2000s when BI architecture and techniques were of principal concern.
6. **Analytics**: the mid-2000s to the present day marked by the increased prominence of analytics.

2.3.1 Data Collection and Analysis

Thus far, the section has considered the overall scope of the historical analysis (i.e. its subject and its boundaries). To conclude, the details of the data collection methods and analysis technique are presented. In respect to the former, the initial starting point is both the timeframes (1910 to the present day) and the disciplines associated with the areas highlighted in figure 14. It is from within this period that sources are drawn, and materials were sourced from a variety of books, journals, magazines and online resources. Such resources are either artefacts published in the time period on which they focus, or historical accounts themselves. Therefore, the approach is, in part at least, best described as a historiographical analysis (e.g. Iggers, 1997), in that the study and the descriptions of the periods of the analysis is of equal import to the actual 'facts' and events.

The analysis technique is somewhat different to a traditional analysis of the kind, in that it seeks to evaluate this history in both horizontal and vertical directions. In respect to the former this is the more standard linear analysis over time, however we are also interested in the verticals of the three areas of technology, quantitative methods and decision making, and how effects have dispersed through the different disciplines we associate with each. These analyses, and the recommendations and conclusions they suggest, are presented in chapter three.

2.4 Job Advert Analysis

As detailed in section 1.4, a significant area of concern is determining the requirements of analytics employers. This has obvious implications for the overall research goal, and assessing the suitability of current provisions in UK universities. Accordingly, this represents RO3, and will be met in part with the job advert analysis of chapter four, the methodology of which is described here.

The data source for this analysis has already been identified; analytics (and related) job adverts (the data extraction and pre-processing steps are detailed in section 4.1). The use of such data is not entirely novel, although has not been employed previously to this specific domain. Sodhi and Son (2010) provide a content analysis of jobs associated with the OR discipline to assess the skills typically required for OR professionals. Other studies in the literature which similarly analyse job adverts, cover areas including information systems (Todd *et al*, 1995; Chao and Shih, 2005), public sector (Redman and Mathews, 1997), and leadership jobs (Den Hartog *et al*, 2007).

However, extracting meaning from text data is not straight forward. In the above studies, the typical solution is to first determine coded taxonomies which can link word counts in the adverts to topics of interest. Whilst such an approach is common, it does present limitations, particularly as the associations between words and topics must be determined a priori. There are two key issues with this. Firstly, that it necessitates a reasonably comprehensive understanding of the topics and themes of interest and may present the same limitations (albeit to potentially a lesser extent) as a questionnaire. Because the relative recency with which analytics has come to prominence, and some of the ambiguities that surround it, it is not possible to achieve this with complete confidence in this case. Secondly, a significant problem occurs with words which may link to multiple topics. In such cases the researcher must resort to fairly basic, and generally unsatisfactory, work-arounds such as counting word frequencies against all applicable topics, ignoring these words all together, or 'splitting' the word count between topics.

For these reasons, and in keeping with the theme of the subject matter, we prefer a more automated approach that can find patterns in the adverts in a more objective fashion. With the advances of processing power in the current age, the options available to researchers to this end have dramatically increased. Consequently, the method adopted in this work draws upon modern techniques in text analytics and natural language processing, which, in combination with the significance of the research question, will deliver a number of important new theoretical and methodological contributions to the literature.

2.4.1 Topic Models

As stated at the start of the section, a key part of the research strategy was not to predefine the topics of interest or the words we associate with them; instead preferring for these to be inferred directly from the data. The family of statistical models used for such a task is known as topic models, of which the most widely used is latent Dirichlet allocation (LDA) introduced in Blei *et al.* (2003). LDA, and the modified versions of the approach developed since, has significant advantages over related methods that predate it, such as singular value decomposition (SVD) and latent semantic analysis (LSA). In SVD and LSA, each word is effectively positioned at a single point in dimensional space and, in other words, cannot deal with polysemy, words having multiple meanings. In comparison, words are assigned to topics on a one-by-one basis such that the same word can be assigned to multiple topics across the corpus. As a toy example, the word "lead" can refer to a wiring in a document discussing electrical engineering, a prospective customer in a document discussing sales, and the element Pb in a document on chemical engineering. This discussion is expanded upon in appendix item F.

LDA has been widely studied and used in many applications. As these methods are less frequent in the OR literature, some brief discussion of the approach will be given. Fundamentally, the model simply regards any document in any corpora as a collection of words that are included based on the topics that are present in the corpus, where each topic is a multinomial over the vocabulary present in the corpus. Whilst all documents share all topics, the proportion of words drawn from any given topic in each document has stochastic variation, as they are random draws from a Dirichlet distribution.

LDA is a hierarchical model containing the known parameters of words and documents, and the latent topics and topic proportions, as demonstrated in figure 15. In the plate notation of the figure, the total set of documents (M) contain collections of words (N), so that $w_{i,j}$ represents the jth word in document i. Each word is generated from one of the topic from the total (K) across the model and is shown as the topic assignment $z_{i,j}$ for the jth word in document i. The process in which the ith document is conceived to be generated, is that each word is drawn from the relevant topic, based upon the word's topic assignments. At a corpus level, α represents the parameter of the Dirichlet prior effecting per-document topic distributions (θ_i), whilst β the parameter of the prior on per-topic word distributions (φ_K). Both the alpha and beta priors (α and β), which effect the sparsity of the document-topic and topic-word distributions, were initialised as symmetrical priors with the value $1/K$ (in this case 0.01).

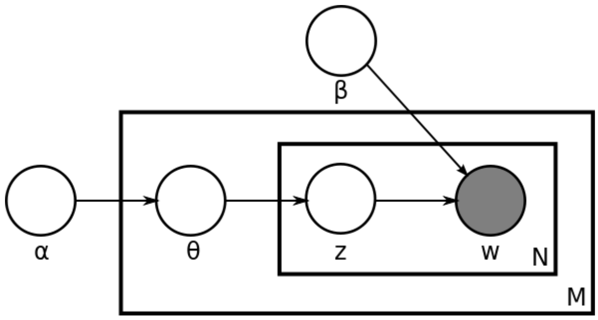


Figure 15 Latent Dirichlet allocation (Source: Blei *et al*, 2003)

LDA, on the assumption that the documents were generated in this fashion, seeks to reverse engineer this process. We can see the output, but not the topics that have generated it. In other words, the goal is to learn, via Bayesian inference, the set of topics and the words associated with each, the topic assignment for each word in the corpus, and the topic distribution of each document. Such an output allows us to qualitatively infer the subject matter of each topic (via its most probable words), and an understanding of differences between documents (job adverts) by contrasting their respective topic propositions, which represents the subjects they discuss (some of which will are likely to correspond to the sort of skills we seek to compare).

Whilst LDA has been shown to be a very effective solution to problems of this kind across a wide range of application areas, when using larger datasets, it can suffer from performance issues due to the complexity of the approach. Although at ≈40,000 documents (see section 4.1 for more details on the data volumes and the split between categories) the dataset is not particularly big by modern standards, it is large enough to seek more computationally efficient methods. One such solution is Online LDA, presented in Hoffman *et al.* (2010). The principal innovation of the approach is to remove the need to pass through the entire dataset with each iteration, by processing a subset of the data (a 'chunk') and, using variational expectation maximisation (e.g. Wainwright and Jordan, 2003), estimating the parameters as if this represented the total corpus. This is done by fixing the topics (λ) and performing an E (expectation) step to find locally optimal values for the other parameters. The next (M) step is to calculate $\tilde{\lambda}$, the optimal value for λ given these parameters. Thereafter, and with each subsequent chunk, the value for λ can be estimated as a combination of $\tilde{\lambda}$ and its previous values (weighted such that old values of λ can be forgotten over time). The final value of λ, after all iterations and passes are complete, can then be used to set the remaining parameters of the model.

These steps, as well as the pre-processing stages detailed in section 4.1, were performed in Python using the packages Gensim (Rehurek and Sojka, 2010), NumPy (van der Walt *et al*, 2011), and the Python Data Analysis Library (PyData, 2012). The model successfully converged, and its outputs were used in many of the subsequent analyses described in the remainder of this section.

2.4.2 Data Analysis

The final part of the methodology is to design the modes of analysis to explicitly meet the objective of the research, establishing the skills requirements of analytics. In this respect, our initial and primary analysis is the topic model itself. Through the distribution of words within topics (i.e. their relative frequency), this provides an overview of the different types of skills involved, and their respective quantities. This not only offers a suggestion of the key topics and requirements of analytics roles, but also can be used to make a comparison with those of OR roles, in reference to those of the other disciplines included in the dataset.

The danger with using this alone, however, would be that this remains somewhat undirected. Importantly we are interested in focusing our comparison on the skills and experience requirements, not, for instance, whether the job requires a work visa or offers a pension. Therefore, the first step is to focus the comparison on the relevant groupings of topics, namely the hard and soft skills and domain requirements identified at the conclusion of section two.

One additional advantage of this approach, is that this allows for a more reasoned selection of the elements to include (much as one would do if using an a priori taxonomical approach (or indeed a questionnaire) as with much of the previous literature). Accordingly, the topics can be analysed as to their relevancy to the task in hand, and those which are outside of our scope (for instance topics related to application processes) can be excluded. The remaining topics, those relevant to our subject area, are then ranked by their average prominence (based upon the averaged θ_i values) within their discipline.

The rankings of the topics can be compared both visually (qualitatively) and then also statistically. For the latter, we employ Spearman's correlation coefficient to return a similarity metric for each discipline in comparison to the analytics subset. If a discipline's rankings are the same as the rankings for analytics, the correlation coefficient would be 1 (perfect correlation). For perfect inverse correlation, the coefficient would be -1, whereas for completely independent rankings the coefficient would be 0. Therefore, the higher the coefficient, the stronger the indication that the job roles of that discipline correlate with those in analytics.

It is worth noting that it would have been possible to use averaged topic distributions as continuous data, either in Spearman's test or the related Pearson's correlation test, instead of reverting to the ranks (effectively treating the results as ordinal data). The reason for this conservatism is concerns about the extent that our topic distributions truly represent ratio-type data. This is not as such a fault in the topic model itself, but moreover the nature of the specific task. The nature of the data source itself necessitates that some of the topics will be irrelevant to the goal of the research (for instance, topics concerning the recruitment process), and the extent to which adverts incorporate content in these areas varies greatly. Accordingly, our topic distributions cannot be considered as true ratio data, and to use this for further analysis is problematic. However, converting to rankings allows a far more like-for-like comparison.

In other words, we can sub-divide the topics of interest into the categories of hard skills, soft skills and domains, and use these to evaluate comparisons between analytics and each discipline. Additionally, in the interest of exploring these issues further, a qualitative analysis of how prominent topics differ across disciplines can be used to understand these differences are manifest. To compare analytics and OR adverts at a more general level, we perform a further tests. Continuing in the vein of the earlier analysis, Spearman's correlation co-efficient can also be used across the whole spread of relevant topics (with, again, those related to more generic aspects of job adverts ignored). This will provide a single metric of comparison to assess the strength of OR's relationship with analytics, in comparison to the relationships of the other disciplines in this study. These analyses were performed in sequence, and are described in chapter four.

2.5 Degree Materials Analysis

The second quantitative analysis, and third instrument overall, is designed to investigate current analytics provisions in higher education. In respect to the research objectives this is decomposed into two parts. Firstly, we seek to identify which degrees in 'traditional' disciplines analytics degrees most closely align with (RO4). Secondly, we seek to identify and evaluate the current content of analytics degrees (RO5). This section describes the methodology employed to meet RO4, and to meet in part (alongside the interviews with academics) RO5.

Analysing degree curricula materials is a task that can be performed in many ways. Firstly, a purely algorithmic approach can be employed, although no examples were found of this in respect to curricula data in the current literature. Whilst this can fail to identify some patterns that are recognisable to the human eye, equally there are other structural patterns that may be more easily found in such an approach. Additionally, these methods can eliminate many of the judgement calls that qualitative methods necessitate, decisions that can lead to potential biases in the results. Finally, a benefit of such approach is the ability to 'scale up'. In qualitative analysis, the time taken is almost directly linear to the number of documents. For algorithmic approaches, most of the time overhead is in the setting up and validation of the model, such that these approaches mean a far greater quantity of materials can be analysed whilst remaining within a reasonable time frame.

Secondly, 'hybrid' methods can be employed where (qualitatively derived) dictionaries of codes are built and applied to degree curricula to identify (quantitative) patterns (e.g. Chu, 2006). This approach has obvious benefits, particularly as it allows the researcher to 'correct' the data by finding synonyms and patterns that are far harder to identify algorithmically, due to the complexities of text data as a whole. For example, a common approach is to create taxonomies of associated terms such that the researcher can record word counts in a topic rather than the words individually. A principal benefit is that, for a knowledgeable reader, the words "masters", "MS" and "MSc" can be regarded as the same, but for a machine this similarity is harder to identify without prior instruction. However, equally there are drawbacks. Whereas in much of the prior research the authors have smaller scope, and therefore can build taxonomies purely on a priori theory, as alluded to in the introduction, we conceive analytics to be a wide-ranging practice encompassing many different traditional disciplines. Similarly, these disciplines are not perfectly partitioned. So, for example, the keyword "data" has linkages with IS, computer science, statistics and OR (among many others). In such a case, we either need to disregard the word entirely or to include it as a count in each of the individual categories. Whilst this can seemingly resolve our problem, the relative counts of a word such as "data" could be very significant in identifying the properties of each disciplines.

Finally, and most commonly in the literature, traditional qualitative methods such as ethnography, cases studies or textual analysis can be used to evaluate the contents of course materials and the teaching methods used (e.g. Stern, 1998). Such methods allow for a deep investigation of the topic, but, as trade-off, the scale of the investigation is likely to be limited (due to time resources) and there is the potential for subjective bias to be introduced.

As this discussion highlights, each approach has its merits, but equally they have their drawbacks. Considering this, our methodology sought to apply all three (to some extent). We firstly performed a large-scale quantitative analysis of the data using machine learning techniques (an algorithmic approach). Thereafter, we performed a closer, 'hybrid' analysis of the results by creating coded 'themes' to detail the modules the degrees offer. In both cases, we used qualitative content analyses to validate and explain the results. Each of these approaches are discussed in sequence, before finally the data collection and transformation processes are detailed.

2.5.1 A Large-Scale Quantitative Analysis of Analytics Degrees

The first (purely algorithmic) analysis is designed to identify the academic disciplines with which existing analytics degrees most closely align. In respect to the scale of the task, and the debate above, this was conceived as a classification problem.

There are many different methods that have been used to algorithmically classify text documents, but one of the most common are support vector classifiers (SVC), introduced in Cortes and Vapnik (1995). Put simply, a SVC works by evaluating some training data (that is some data which already contains class labels), and identifying a hyperplane (or hyperplanes in multi-dimensional space) that can separate one class from another. If a dataset were plotted, a well-fitted hyperplane would be the line separating the two classes (such as one class sits one side of the line and the second class the other) which maximises the distance to the nearest data point. Once the position of the hyperplane has been established on the training data, the model can be used to classify new unlabelled data based on which side of the hyperplane it is situated.

SVC is highly suited to our task, and to working with text data in general. Joachims (1998) describes several reasons for this including:

1. **High dimensionality:** as our analysis of job adverts is based on the words within them, effectively the number of dimensions is equal to the size of the vocabulary, which tends to be large (over 10,000 in this case).
2. **Feature relevancy:** as above, text data tends to have a high number of features. In most classification algorithms (such as decision trees or naïve Bayes) relatively few features are used to perform the classification, either determined algorithmically or through feature engineering. While in practice these approaches can still be effective, as experimentally demonstrated in Joachims (1998) even using the features (words) with the lowest information gain to build a classifier perform better than a random model, indicating that these features are not redundant and can help an algorithm classify more accurately.

3. **Data sparsity:** although there are typically a large number of features/dimensions in text data, when represented as a document-term matrix (a M x N matrix of word counts with one row for each document in the corpus, and one column for each word in the vocabulary – discussed in more detail in appendix item F) most of the entries will be zero. This is a common problem in text analytics, but one which SVC, as an "additive" model (Kivinen *et al,* 1997), is suited to.

SVC can be linear or non-linear (by projecting the data into a transformed feature space using the 'Kernel trick' (Aizerman *et al*, 1964)). Using non-linear kernels such as the polynomial has been shown particularly effective in natural language tasks (e.g. Renders, 2004), however, linear SVC has also been successful (e.g. Fan *et al*, 2008). Therefore, both kernels were tested for performance. Another concern is the approach used to adapt SVC to a multi-class problem such as this, as fundamentally SVCs, as detailed above, are binary classifiers (i.e. they can only separate two classes not multiple). The method selected was a one-vs-rest approach, whereby individual binary classifiers (in class / not in class) are built for each category.

Although SVC has been shown in the literature to have been successful with such tasks, this is not to say that their performance cannot be improved. One of the most common meta-algorithms for such a task is bootstrap aggregation (bagging), whereby bootstrap sampling (sampling with replacement) is used to create multiple classifiers, whose results are aggregated (as 'votes') to determine class assignment. Breiman (1996, p. 124), in his initial description, argues bagging offers the researcher "improvement [...] for unstable procedures". Due to the issues associated with text data and with internet sources, there is sufficient reason to believe this may qualify as an "unstable procedure".

Obviously the principal of "no free lunch theorem" (Wolpert 1996), that no algorithm will be superior across all types of problem, would apply, and there is no claim made that there has been an exhaustive search of potential solutions. However, for these reasons, SVC was considered a good fit for the problem presented here, particularly with the addition of bagging to help improve accuracy and prevent over-fitting.

2.5.2 Weighted Module Analysis

The second (hybrid) approach, designed to identifying the skills, subjects and techniques taught within analytics degrees, combines both qualitative and quantitative elements. Firstly, the modules incorporated in each program were extracted from the degree materials. Secondly, the modules drawn from the disciplines associated with analytics (i.e. all the degrees except those labelled as "analytics") were coded based on their principal themes. The codes were created inductively, and iteratively updated during the progress through the corpus. The counts of the occurrence of each code (one per module) were retained. This allows a comparison of the influence of the different disciplines in the dataset of analytics degrees, by identifying the module topics most closely linked to each discipline and computing an "association score" for each.

One concern in this process is the level of granularity to apply in this coding. For example, a module in C# programming could be coded as specifically this title, as C programming (i.e. the whole family of coding languages), programming as a whole, or even a more generic code such as computing. In general, the approach used was to seek a reasonably granular level of detail in the codes deemed to be most important to analytics (drawing on the existing literature), so we chose the code label C programming in the specific example given here, and then broader codes for the topics that are less directly related (for instance, both "brand marketing" and "international marketing" were labelled as "marketing"). The second main concern was around modules that covered topics which incorporated more than one code. For example, during the analysis separate codes for machine learning and data mining were created. To classify a module titled "machine learning and data mining", we simply allocated a count of 0.5 to each code.

Although some insight can be gained from simply analysing the overall counts, many topics will be recurrent across multiple disciplines, and due to the different quantity of materials extracted in respect to each discipline (discussed in section 2.5.3), counts alone may not represent the relative frequency across disciplines. Therefore, we sought to introduce a scoring system that accounts for the relative importance of each code (module topic) in comparison to other disciplines; in other words, the codes which had the most discriminatory power in characterising each of the degree types.

To do so, borrowing from the widely used χ^2 test, we compute an 'expected' count for each term in each discipline; that is a calculation of the frequency one would expect if the term was distributed proportionally across the different disciplines. We can then compare the actual frequency of the topic in each discipline to its expected count, such that if the actual exceeds the expected, we conclude the topic is important to the discipline. However, we also need to control for the fact that some disciplines have a greater number of topic codes (primarily due to the disparity in sizes of datasets). Therefore, we finally take the amount that the actual exceeds the expected as a proportion of the total quantity of observed occurrences to give what we describe as a Module Topic Weighting (*MTW*) for each term and each discipline. This approach can be written algebraically, with *MOF* as the Module Observed Frequency and *MEF* as the Module Expected Frequency, as:

(1) $MTW = \frac{MOF-MEF}{MOF}$

Where (with *TMD* as Total Modules in Discipline; *TFT* as Total Frequency of Topic across all disciplines; and *TM* as Total Modules):

(2) $MEF = \frac{TMD}{TM} \cdot \frac{TFT}{TM} \cdot TM$

There are two significant differences between these equations and those used in the χ^2 test. Firstly, in contrast to equation (1), in the χ^2 test one would square the difference between observed and expected and divide by the expected. In this instance, we do not need to square the difference (as we are only interested in positive results) and we measure as a proportion of the observed rather than the expected as we are more interested in the frequency of the term than in the scale of the difference. Secondly, we do not seek to compute a test statistic to compare to the χ^2 distribution. In essence, this is an omission that is forced upon us. The common rule of thumb, that expected counts should be at minimum of five, would be violated in the vast majority of cases since the number of variables (topics) exceeds the number of cases (degrees). Whilst this clearly diminishes the claims we may make of the analysis' validity, as we are not able to apply significance testing, ultimately, we treat the results of this analysis as an indicator rather than statement of fact, and represents a part of a series of methods upon which we reach our conclusions. However, to ensure that the figures are not distorted by outliers, we only retain MTW scores for terms that occur more than twice in each discipline, and represent greater than 1% of the total topics of that discipline.

Having created MTW scores for each topic in each module list of the degrees associated with the disciplines listed (but excluding analytics), this now gives a quantitative basis on which we can assess the relative importance of different degrees in the modules offered in our analytics degrees. To do so we applied the same coding structure to the module lists of the analytics degrees and again retained the counts. We then multiply the frequency of that topic across analytics degrees by the MTW score associated with each discipline and finally summing these by discipline. As such, we ultimately produce a final score for each discipline (which is a combination of the frequency of the topics in analytics degree modules and the relative importance of these topics to degrees in that discipline) from which we can compare the relative influence of the discipline on analytics degrees. Each of these analyses, bagged SVM classification and the weighted module analysis, were performed in sequence and are presented in chapter four.

2.6 Interviews (Employers)

The last two research instruments were both in the form of interviews, across two separate groups of interest. The first of these were with employers, and potential employers, of the graduates of master's degrees in analytics and (to a lesser extent) OR. The insights that can be gleamed from these interviews give a deeper perspective on the skills required of such graduates, and can be used to complement the job advert analysis of chapter four (and thus meeting RO3). The remainder of the section details the methodology employed to this end, focus on the participants, interview process and analysis methods.

2.6.1 Interview Participants (Employers)

The recruitment strategy for participants was effectively convenience and snowball sampling. However, effort was made to loosely fit certain target quotas (i.e. no specific target was given, but an attempt was made to find balance between certain criteria). This led to the specific recruitment of some of the participants when target groups were lacking. A concern was to see representation across four categories (employers; analytics software vendors; consultants; and recruitment specialists in the area), as well as to cover a variety of industries. This was designed to ensure a wide enough variety was represented, and that no bias was introduced by focusing too heavily on specific sub-populations. In total, there were 29 participants (including an initial pilot interview which was included as the questions did not require any significant alterations). The distribution of respondents across categories and industries is shown in table 15.

Table 13 Interview participants by category and industry

Categories	*n*
Employer	10
Consultant	10
Vendor	7
Recruitment	2

Industries	*n*
Software	7
Management consultancy	6
Public sector	5
Marketing	3
Utilities	3
Recruitment	2
Travel	2

Whilst the split shown in table 15 is not quite perfect, it is considered suitably diverse to ensure no bias is introduced, and that a variety of perspectives are included. Also, a suitable balance is found between private and public-sector employers, as well as covering three prominent domains in analytics (marketing, utilities and travel).

2.6.2 Interview Strategy (Employers)

The interviews were conducted in a semi-structured approach, and recorded and transcribed post haste. Template questions were included, but adapted slightly dependent on the category of participant (for instance, for direct employers the questions regard skills required of new recruits, for consultants or software vendors the same questions, generally, relate to the demands of clients in the same areas). The interview questions in full are included as appendix items B and C, whilst a summary of the main topics shown in table 16.

Table 14 Interview questions employers (abridged)

Background	Q1	A little information about yourself / your business
Definition & Use of Analytics	Q2	How do you define analytics / data science in terms of your business?
	Q3	What data storage and management tools do you use (e.g. Hadoop / MySQL / Oracle / etc.)?
	Q4	Which analysis tools do you use (e.g. SAS / R / SPSS / etc.)?
	Q5	Which, if any, programming languages do you use?
	Q6	Do you use any data visualisation tools? What are they used for?
	Q7	How is analytics / data science managed in your business?
Recruitment & Education	Q8	Have you been recruting or are currently recruting for analytics staff?
	Q9	Are you involved in any university initiatives to support training of new graduates (e.g. guest lectures / case study materials / etc.)?
	Q10	What do you think are the key skills, capabilities and experience degree courses should offer graduates working in analytics-type roles?
Future Developments	Q11	How do you think analytics will develop next in your business or your client's businesses?
	Q12	How do you think analytics will develop in general?

As indicated, the questions are broadly separated into three main categories: those concerning the definition and use of analytics within their organisation; the recruitment of analytics staff and the skills they feel candidates should bring; and finally, how they perceive analytics will develop in the coming years. Within these there are question regarding technology aspects, quantitative aspects, and the use of decision support tools (aligned with the taxonomy of figure 14, section 2.3). The additional sub-questions (shown in the appendix) also incorporate issues regarding the spread of analytics expertise between functions, internal training, and the balance between hard and soft skills required of recruits. The interviews were conducted over a period of four months.

2.6.3 Interview Analysis (Employers)

In consideration of the philosophical positioning of the work, described as 'centre-right' pragmatism – employing MMR in a loosely realist framework, certain analysis methods become more appropriate than others. For instance, a truly grounded-theory type approach would not fit comfortably with the sequential nature of the research design. Accordingly, the main analysis method employed is template analysis, an approach noted by King (2004, p 256-257) for its application "within a range of epistemological positions [...] template analysis is, on the whole a more flexible technique [than grounded theory] with fewer specified procedures, permitting researchers to tailor it to match their own requirements".

The basic premise is of thematic coding, with similarities with both grounded theory and, more significantly, Interpretative Phenomenological Analysis (IPA). As with IPA, the main process is the construction of codes, which are then grouped thematically to create a hierarchical structure of themes within the text (the template). The template is first constructed on a subset of the cases, and then used to analyse the remainder (with the template being updated as required to fit the thematic content of the transcripts). Key aspects of the analysis are an ability to compare the template between cases (which will be a key part of the analysis of academic responses), and the ability to chart the evolution of the template from its initial form to the final version – effectively the development of knowledge about the phenomenon in question as the analysis progresses. In respect to this latter point, in that template analysis allows for *a priori* coding, a key difference from IPA; effectively we can measure the development of our understanding of employer requirements from the initial insights drawn from the literature and (quantitative) job advert analysis, through to the final template.

As an additional extension, matrix analysis (Miles and Huberman, 1994) can be applied to subsets of the respondents. Matrix analysis, which is a stand-alone technique but one which also can fit well with the results of a template analysis (e.g. Nadin and Cassell, 2004). The basic premise is to build matrices of the data where the rows represent participants and the columns key themes of interest (in this case drawn from the template analysis). The matrix approach therefore allows for the comparison of participants across key topics within the analysis. Whilst a full, stand-alone matrix analysis would feature all participants in the study, in this research specific participants are selected to represent sub-groups that are found in the template analysis based on their use of analytics and *a priori* categorisation (see table 15, section 2.6.1). The results of both analyses are presented in chapter six.

2.7 Interviews (Academic)

The final instrument used in the research is again semi-structured interviews, however, this time focused on the other population of interest, academics involved in analytics courses. The analysis of these interviews is designed to complement the quantitative curricula analysis (chapter five) to create an understanding of current education provisions for analytics (RO5), as well as identify the potential barriers and obstacles that may impact future academic development (RO6). As before, the section will detail participant recruitment, interview strategy and analysis methods.

2.7.1 Interview Participants (Academic)

As before, participant recruitment was via convenience and snowball sampling. There is obviously a big gap between the size of the populations, with those involved in analytics/OR higher education as far fewer than potential employers. Although not necessary a design choice, with generalisation not really a consideration with such an instrument, there is an obvious impact on the availability of respondents and an expectation that saturation can be reached more quickly. Accordingly, a far smaller sample was used for this part of the research with 11 separate interviews conducted, at 8 institutions (although one interview was conducted with 3 participants (all involved in the same master's course) simultaneously – meaning 13 participants in total).

Another difference is that the categories and industries used to stratify the participants in the former group (employers) do not apply here. However, we did seek to maintain some regional balance to avoid potential biases this may introduce. The geographic distribution of these participants is shown in table 17.

Table 15 Geographical split (institutions) of interview participants

Region	*n*
South East	1
Midlands	2
North West	2
North East/Yorkshire	2*
Northern Ireland/Ireland	2
Scotland	2

* 1 interview was with 3 participants

There are additional potential concerns and biases. Ideally, we sought participants who were familiar with both analytics and OR, which does have implications on a potential bias towards OR perspectives of analytics. In total 9 of the 13 had significant OR training and background. This also had implications, seemingly, on the schools with which respondents were based. 12 of the 13 were based in business schools and another a computing school. Whilst this is a concern, it was not easily addressed so remains as a potential limitation, and something that needs to be considered when interpreting these results.

2.7.2 Interview Strategy (Academic)

The overall interview strategy followed the same general principals as described in 2.6. There was not the same requirement for multiple question sets for this group, however, some adaption occurred. A summary list of questions is shown in table 18, with the full version include as appendix item C. Several of the questions match those of the previous interviews (table 16, section 2.6.2). 1-9 match fairly closely to others in the first set. However, additional questions are posed regarding teaching methods and potential barriers (10-14) that are obviously not relevant to employers.

Table 16 Interview questions academic (abridged)

Background	Q1	A little information about yourself and your school/university?
Definition & Use of Analytics	Q2	What differences do you see between analytics and OR (or other subjects)?
	Q3	What academic traditions do you think analytics draws from?
	Q4	Do you offer analytics courses?
Core Skills	Q5	What are the core skills that you think analytics courses should teach?
	Q6	What are the core skills that you think OR courses should teach?
	Q7	To what extent do you think that core analytics skills are delivered in OR degrees (or other subjects)? What is missing?
	Q8	Are there programming languages that should be taught in analytics or OR courses? Do they differ?
	Q9	Is there specific software?
Pedagogy	Q10	What forms of datasets should be used - and where can they be sourced? Is there a difference between OR and analytics?
	Q11	What types of problems and exercises should be presented? Is there a difference between OR and analytics in this respect?
	Q12	What value do you place on internships or consultancy projects?
Barriers	Q13	What barriers do you see that complicate the creation of analytics degrees?
Future Developments	Q14	To what extent do you think that universities need to adapt to current business trends and how much do they need to maintain the academic traditions of disciplines?
	Q15	How do you think analytics and OR degrees will develop in general?

2.7.3 Interview Analysis (Academic)

To analyse the interviews, again template analysis is employed. To make comparison between employer and academic groups, initially the final template of the former analysis is used and adapted accordingly. The net result will be elements that are not retained (much of which will be relevant to employers but not to academics) and elements retained (providing interesting comparison points). In other words, comparison with the elements of the template retained can be used to meet RO5, whilst the elements added to the template are used to meet RO6.

2.8 Summary

In summary, this chapter has detailed the methodology of the research. The underlying philosophy has been presented, described as a 'centre-right' interpretation of pragmatism. Additionally, the interaction between the different research instruments has been discussed, as well as the limitations of the approach. Finally, each of the five research instruments have been outlined, with their respective data collection and analysis methods. With their respective methodologies explained, the theses will continue by presenting each analysis in sequence, beginning with the historical analysis described in section 2.3.

3 Historical Analysis: From Taylorism to Terabytes

In earlier discussion, two significant inconsistencies are identified in the prior literature. Firstly, there is a general lack of clarity as to precisely how analytics relates to OR and practices such as business intelligence (BI). This represents the first objective of the research (RO1), and of this chapter. Secondly, the computational literature review highlighted a relative absence of academic research into analytics from within the OR community. Therefore, the second objective of both the chapter and the research (RO2) is to determine a research agenda for OR that accounts for the growth of analytics and its specific characteristics and concerns.

Accordingly, the rest of the chapter is arranged as follows. We begin by presenting the six periods of the history described in 2.3 (and relisted below), summarising the key events and movements. Secondly, an analysis of these results is presented, addressing RO1. Finally, the implications for the OR discipline are considered, and a research agenda presented (addressing RO2).

Author's note: the contents of this chapter have been published in Mortenson et al. (2015). This work was written firstly for the thesis, but subsequently packaged for publication. In regards to author contribution, this followed the same process as the remainder of the thesis, with the additional authors taking primarily a supervisory role in terms of concept development, advice and editing.

3.1 Analysis of the Dianoetic Management Paradigm

The methodology presented in section 2.3 describes the source materials used in this analysis, written historical accounts and other sources related to disciplines related to the paradigm, as well as the time frame under investigation (c1910 to the present day). However, as described, we sub-divided this time frame into six periods, none of which constitute paradigms of in and of themselves, but nevertheless have specific characteristics. These periods are shown below in table 19, along with some of the key events we associate with each, divided into the categories of *technology*, *quantitative methods* and *decision making*, from the Venn diagram of figure 14 (section 2.3).

Table 17 Selected events in the dianoetic management paradigm

		Technology	Quantitative Methods	Decision Making
First Period: *Scientific Management*		**c1913:** The Ford Model I began production using its influential assembly lines **1914:** The end of the *Technological Revolution* **1941:** The first digital computer, the Z1, released	**1935:** Publication of Fisher's *The Design of Experiments* **1938:** First discussions of "OR" (Kirby, 2003, p 71) **1939:** Development of cluster analysis	**1912:** The principles of Gestalt visual perception devised (Wagemans *et al*, 2012) **1921:** Launch of the Cambridge Psychological Laboratory designed to distribute study results amongst industry
	1940s			
Second Period: *The Scientific Method*		**1945:** Design of the *von Neumann Architecture*, structures still used today **1952:** The UNIVAC computer predicts the US presidential election **1957:** FORTRAN programming language devised	**1947:** Linear programming developed **c1947:** OR/MS methods used to help rebuild UK industry (Kirby, 2003, pp 190-205)	**1946:** Formation of the *Ergonomics Society* **1947:** Publication of Simon's *Administrative Behavior* **c1959:** The development of an air defence system with the first graphical user interface (Gurer, 2002)
	1960s			
Third Period: *Management Information Systems*		**c1963:** The development of microchips **1964:** Release of the IBM System/360 **c1970:** E. F. Codd conceptualises the first relational databases (Date, 2000)	**c1963:** Geography's *Quantitative Revolution*, demonstrating the growth of quantitative methods across various academic disciplines (Burton, 1963) **1964:** Lancaster University launches the first UK master's degree in OR/MS	**1962:** The *Myers Briggs Type Indicator* published, used to better understand decision maker types & needs **c1962:** Behavioural science grows in influence, particularly in consumer research (Kardes *et al*, 2010)
	1970s			
Fourth Period: *Decision Support Systems*		**c1972:** Personal computers are popularised in business (Ceruzzi, 1999, pp 207-241) **c1972:** TCP/IP internet protocols introduced **1973:** IBM 3660 Supermarket System released introducing barcode scanners	**c1975:** 'S' statistical language & Matlab are launched. SPSS & SAS grow in popularity (Wegman et al, 1997) **1979:** Development of the ID3 decision tree algorithm (the predecessor of C4.5)	**1979:** Research into decision making needs of CEOs leads to the design of Executive Information Systems (Rockart, 1979) **c1981:** Development of the 'soft systems' approach to decision making
	1980s			
Fifth Period: *Business Intelligence*		**1988:** The conceptualisation of data warehouse architecture (Devlin and Murphy, 1988) **1989:** Launch of the world-wide-web **1993:** IBM Simon released, the first smartphone (Lewis, 1996)	**c1988:** The first significant research into agent based modelling (Samuelson, 2000) **1989:** The term 'data mining' introduced (He, 2009) **c1996:** General Electric introduces Six Sigma into its operations (Henderson and Evans, 2000)	**1992:** Development of *balanced scorecards* (Kaplan and Norton, 1992) **1999:** The release of *Grammar of Graphics*, a set of rules for data visualisation **2000:** Popularisation of dashboards (Marcus, 2006)
	2000s			
Sixth Period: *Analytics*		**2004:** A paper from Google's Dean and Ghemawat details MapReduce, a programming paradigm for big data **2004:** Launch of Facebook (Twitter launches in 2006) **2007:** Development of NoSQL databases (Driscoll, 2012)	**2001:** The release of the Natural Language Toolkit in Python, helping popularise text mining **2008:** Publication of Anderson's *The End of Theory* **2010:** The first Kaggle competition (Yang, 2010)	**2005:** eBay buy shopping.com, illustrating the importance of recommendation agents (Xiao and Benbasat, 2007) **2013:** Visualisation software vendor Tableau is valued at $2bil after two days on the stock exchange (Cook, 2013)

In consideration of table 19, there are of course many innovations and events that could have been included, and almost any knowledgeable reader could point to key aspects that have been omitted. The table is intended to provide some examples and is in no way an exhaustive list. Some other key events are detailed in the later discussions of this chapter, to help assuage some of the sense of arbitrariness that such a table will inevitably bring.

Each of these periods are discussed in sequence, beginning with the *Scientific Management* period, the first of this history.

3.1.1 The First Period: Scientific Management (1910 – 1945)

As stated, the proposed start of the first period, and overall dianoetic management paradigm, is circa 1910; not only marked by the publication of Scientific Management but also the closing stages of the Second Industrial Revolution (also referred to as the Technological Revolution). Smil (2005, p 8) argues these "widespread and truly revolutionary innovations not only changed the course of the innovating societies but were also translated into profound global impacts". These global impacts can in part be seen in the changing approaches to management of this paradigm. New technologies begot new products, services and industries, but also new methodologies that impacted upon not only physical labour and methods of production, but also management approaches (e.g. Fordism). The principles, methods and philosophies of process management developed by Taylor, Ford and others were to have sustained influence, much as the technologies themselves influenced them.

World War Two was a period of significant innovation, most famously in Bletchley Park where Colossus, the first programmable digital computer, and decryption machines (such as Alan Turing's Bombe) were created (Randell, 1972; Flowers, 1983). Similarly, the work of Edward Tizard, Patrick Blackett and the Aeronautical Research Committee, arguably the originators of the OR discipline, played a noteworthy influence on Britain's war effort (Ormerod, 1999; Kirby, 2003), as well as many of the *quantitative methods* of the dianoetic management paradigm. Whilst the work carried out in Bletchley Park became so widely acknowledged it ultimately become almost folklore, less celebrated advancements were occurring around the world. In Germany Konrad Zuse created the Z1 (the first digital computer) predating Colossus by two years (Giloi, 1998), whilst mathematics became increasingly important in military operations of the US (Rees, 1980) and Canada (Laporte, 2008). A well-known example of this, was the application of Monte Carlo simulation, based primarily on the work of Stanislaw Ulam, was widely used in the Manhattan project (e.g. Metropolis, 1989).

Elsewhere, and outside of the war effort, there were other significant developments, particularly in the application of quantitative methods to business and management problems. A notable example was the interest and work around demand curves, such as Working (1927), of high relevance in a world where the innovations of Ford and his contemporaries meant consumer goods could be mass produced faster than there may be demand for. Such work was highly influential in the development of the field of econometrics and quantitative economic research, furthered by the establishment of the Cowles Commission for Research in Economics in 1932.

In summary, this period is when the innovations of the Technological Revolution began to impact on managerial theory and process. Similarly, the period demonstrates the domino-effect of interactions between different disciplines and society: new technological innovations led to changes in working lives and practices, which in turn inspired new approaches to management and the new paradigm.

3.1.2 The Second Period: The Scientific Method (1945 – mid-1960s)

Following the conclusion of the war, the pioneers of the nascent computer technologies and the OR discipline sought new applications for their tools and methodologies. Moreover, a recognition of the potential cost savings each offered was not lost on the cash-strapped governments of Europe and North America. In the UK, the newly elected Labour government, seeking to increase the size of the public sector, engaged Blackett and colleagues to utilise OR in a succession of new industries such as steel and coal mining (Ormerod, 1999; Kirby, 2003). Although mostly regarded in the UK as a smorgasbord of techniques from a variety of approaches, in the US a formulisation of the methodologies occurred and by the 1960s many of OR's principal techniques were established (Kirby, 2003).

As OR grew in popularity (and application), many of the other quantitative disciplines we associate with the paradigm were also gaining prominence and new methods and innovations were introduced. In 1952, statistician Robert Goodell Brown published *Exponential Smoothing for Predicting Demand*, a key milestone in forecasting and econometrics. In medical research, the Kaplan-Meier estimator was published (Kaplan and Meier, 1958), used to assess the best treatment options for patients, an algorithm considered to have "saved millions of lives" (Champkin, 2014).

The second period also saw an explosion of innovation in computing, what Ceruzzi (1999, p 13) describes as the "advent of commercial computing". The list of innovations in the period include the von Neumann architecture (the division of processing and storage memory), the conceptualisation of FORTRAN and COBAL (the first higher-level programming languages), core memory, and the UNIVAC computer (Pugh, 1984; Aspray, 1990). Arguably it was the latter of these which had the greatest public impact by successfully predicting the 1952 US presidential election (Ceruzzi, 1999). The reaction to this was a major public relations coup for the burgeoning computing industry. Indeed, there are many parallels with the reporting of Nate Silver successfully predicting the 2012 election, and the positive attention it has brought to analytics (e.g. Thaler, 2012).

Developments in *decision making* were more limited, though the period did see the formalisation of the disciplines of behavioural science and ergonomics (Senn, 1966; Waterson, 2011). However, the more significant aspects of the period were in the commercial applications of computers and OR, capitalising upon the appetite for a more scientific methodology to business and decision making by demonstrating the actual benefits this can bring.

3.1.3 The Third Period: Management Information Systems (mid-1960s – mid-1970s)

Whilst the computers of the previous period had demonstrated the potential value of such machines in business, their actual dispersion was far more limited. For example, only 19 UNIVAC computers, the most famous of the period, were sold between 1951- and 1954, in what was effectively the machine's heyday (Cerruzi, 1999). It was not until the mid-1960s that computers became accessible to many more businesses, in particular, IBM's System/360, so named due to its targeting of "the full circle of customers, from business to science" Ceruzzi (1999, p 144). Alongside mainframe computers, the period saw the introduction of mini-computers where new efficiencies in storage and logic, combined with a low retail price, generated significant sales across many industries (Ceruzzi, 1999).

The growth in computing had strong influence on the application of these methods. One specific example is the development of the RASCEL computer, designed to implement stochastic methods which until this point were too time and resource consuming for practical application in business (Esch, 1969). Indeed, many of the OR practices such as simulation were particularly boosted by the advent of the computer programs and increased processing power of the age (Ormerod, 1999). However, inasmuch as the period is characterised as being one where technology and quantitative methods came together, it is also worth noting that this is not one-way traffic. In fact, one of the most notable developments of this time, the Box-Jenkins method for time-series forecasting (named after its creators George Box and Gwilym Jenkins, 1968), in many ways moved in an opposite direction. The Box-Jenkins method created an alternative to the now often fully automated methods for forecasting (run entirely on a computer without need for human interpretation), to a semi-automated method which required human input, but better suited many real world time-series problems (Newbold, 1975).

The period saw many developments in academia, with the inception of the University of Minnesota's influential MIS department and the first UK OR master's degree at Lancaster University. The *decision making* aspect of the paradigm also came to prominence, with research conducted at the Carnegie Institute of Technology and MIT, and publications from Simon (1965) and Anthony (1965) particularly influential (Power, 2007). Alongside this more general work into the interface between 'man-and-machine', notable research was published by Scott Morton (1969) and Ferguson and Jones (1969) into how practical system can be devised that would better support decision making.

It was this work (in the main) that provided the stimulus for the transition from this period into the next. Whilst these significant developments in both hardware and software made information systems and data far more pervasive and integrated into businesses, there was still a gap between the potential of the systems and their realised value to quantitative analysts and decision makers. It was attempts to address this gap that provided the catalyst for the start of the next period.

3.1.4 The Fourth Period: Decision Support Systems (mid-1970s – late-1980s)

As discussed in the previous section the fourth period was characterised by a desire to increase the usability of MIS and to further integrate computers into business processes and decision making. This was manifest in three new applications of computing technologies: enterprise resource planning (ERP) systems, expert systems and decision support systems (DSS). The first of these, ERP systems, are arguably less aligned to this history that the latter two, though that is not to say they were not, and indeed still are, key tools for business management. While such systems are not necessarily used to support quantitative analyses, they shared the goal of better connecting organisations and sharing information. Indeed, in an article discussing the history of the software, Jacobs and Weston (2007) argue the motivation behind ERP was "a need for stronger integration between the functional enterprise silos".

Whilst both expert systems and DSS had essentially the same goal, to assist decision makers and improve the efficacy of their decisions, how they sought to achieve this was fundamentally different. Expert systems sought to guide the user to a suggested action, dependent on the specific circumstances of the situation, whilst DSS provided more general decision support, displaying the relevant data or model results to do this (Nelson Ford, 1985). Critically, however, another similarity between the two was that both sought to combine the three categories of the paradigm. Computer technologies underpinned the systems; *quantitative methods* were used in the algorithms and models which analysed the data; and finally, graphical user interfaces (GUIs), influenced by the growing work in disciplines associated with *decision making*, were designed to maximise accessibility and the influence of the systems.

As such, this movement can be characterised as a convergent period, whereby developments in *technology*, *quantitative methods*, and *decision making* were sought to be consolidated into single systems, maximising the impact of each. As an example, by the end of the period many of the leading OR groups began to publish computing-related journals: The Operations Research Society of America (ORSA) with the Journal of Computing; The Institute of Management Sciences (TIMS) with Information Systems Research; and The OR Society began publishing the European Journal of Information Systems. The period also saw the emergence of human-computer interaction (HCI) as both a term and a specific area of academic research, emphasising the overlap of *technology* and *decision making* in the paradigm (e.g. Card *et al*, 1983).

However, that is not to say that all commentators were entirely united on the subject. Echoing the earlier discussion about distinctions between BI and analytics, controversies occurred as to whether DSS was a subset of MIS (Davis, 1982), its evolution, or "just another buzzword to justify the next round of visits from the vendors" (Sprague, 1980, p 1). In parts of the OR community the period too saw disagreement about the influence of the 'softer' side of the paradigm (*decision making*) on its methods and models. Firstly, the period saw the emergence of MCDA, and related approaches such as Analytic Hierarchy Process (Saaty, 1980), methods that framed problems as a combination of "a set of objectively defined alternatives and a set of subjectively defined criteria" (Buchanan *et al*. 1998, p 334). Elsewhere, attempts were made to create solutions to "wicked" problems (Churchman, 1967) problems which are harder to structure and define due to conflicting perspectives amongst relevant stakeholders. This led to the development of the soft systems methodology (Checkland, 1981) and strategic options development and analysis (Ackerman and Eden, 2010). However, these methods were more qualitative in their approach, leading to some degree of polarisation in the OR community as to whether such "soft" methods were appropriate to the discipline; what Dando and Bennett (1981, p 91) would describe as a "Kuhnian crisis".

While such "softer" developments may fit the narrative proposed for the period, that is not to say that "hard" quantitative approaches were entirely absent from this time-period. Several key innovations were developed in this time, including statistical method such as the widely-used Box-Whisker plot (developed by John Tukey, 1977) and bootstrapping (developed by Bradley Efron (1979), and specifically described as an example of the use of modern computers in statistical theory), or data envelopment analysis in OR (Charnes *et al,* 1978).

In summary, perhaps the most significant contribution of the period was to highlight the growing levels of interconnectivity and interdependency across the paradigm. Firstly, this can be considered a conscious effort by researchers such as Keen and Scott-Morton to unify such disciplines, but also this is visible in the ripple effects that the growing influence *decision making* disciplines had on both *technological* and *quantitative* disciplines.

3.1.5 The Fifth Period: Business Intelligence (late-1980s – mid-2000s)

One of the main catalysts for the start of the fifth period came from an unexpected source: the supermarkets. Whilst barcode scanners had been first introduced in 1973, there had been a relatively slow uptake in US supermarkets. However, by 1985, 29% had installed the technology (Basker, 2012). A by-product was the availability of vast amounts of transactional data for retailers and brands. In particular, US consumer goods giant Proctor & Gamble, in conjunction with Metaphor Computer Systems, were instrumental in demonstrating the value and the methodology of a new form of architecture (Nylund, 1999). The architecture amalgamated existing DSS, databases, market research, and the transactional data collected at the supermarket tills into new data warehouses.

However, increased data volumes not only presented technological challenges, but also stimulated demand for new quantitative and decision making approaches. The discipline of data mining attained both credibility and momentum, primarily due to the challenges created by the comparatively large datasets that became available in the period. Through combining statistics, SQL, and machine learning, data mining grew to offer credible and effective solutions. Similarly, the period saw the development of dashboards. Whilst these were still essentially GUIs, in contrast with the first DSS, these dashboards were prepopulated with key performance indicators (KPIs) designed to speedily convey the critical measures of business performance (Few, 2006). The use of such metrics as management tools had become popularised by Kaplan and Norton's (1992) balanced scorecard. Through a combination of this framework and dashboard technologies, the period created something of a culture of management by metrics whereby KPIs, were used, in some organisations, to determine everything from staff bonuses to strategic and operational decision making (Beatham *et al*, 2004).

In summary, the BI period was most notable for the introduction of new architectures and procedures which made the storage, management and delivery of data within the organisation far more efficient and consistent. Much of the catalyst for this was the significant increases of data available at the start of the period. However, the development of the internet during the period, would, by the start of the next, produce an influx of data the scale of which was incomparable.

3.1.6 The Sixth Period: Analytics (mid-2000s – present day)

Analytics as a term can be traced back to Aristotle and his work on deductive reasoning (Malink, 2012). In business, the term is first used around 2000 (e.g. Whiting, 2000), and in the context of BI software. The first academic article identified in this research explicitly discussing the subject is from 2002. In this article Kohavi *et al.* (2002) highlight five key drivers: “verticalization” (the creation of bespoke software for more industries); increased accessibility of models to different business users; analysis tools better integrated into information systems; cross-functional usage in different business ‘silos’; and uses in performance management. However, they also specifically acknowledge another major factor, the growing amounts of data. Similarly, Davenport and Harris (2007, p 11) cite key catalysts as the fact there is far more business data available than ever before and “a new generation of technically literate executives – the first to grow up with computers”.

The growth in data, a key factor as indicated above, is mostly attributable to the ubiquitousness of the internet in the new period. This has had significant ramifications for businesses in terms of data-availability including data from competitors, customers, the general public (through social networks and user-generated content), machines and products (the ‘internet of things’), and in the business itself. This data is of such scale as to limit the application of BI architecture and relational databases (Stonebreaker *et al*, 2007), creating a demand for new technologies and architectures. Most notable is perhaps Hadoop, a distributed file system (DFS), designed to store, process and analyse such data but also includes NoSQL and NewSQL databases (Cattell, 2010); the proliferation of cloud computing; and API-streams from data-rich sites such as Facebook and Twitter. In short, there has been a completely new ecosystem of businesses, technologies and cottage industries built to tackle the challenges of big data (see Feinleib (2012) for a visual representation of this ecosystem).

As mentioned, data scale and complexity has also created challenges for quantitative analysts, and indeed ideological debates. The prevalence of unstructured data (mostly from online sources) has led to further developments in text mining, network analysis and natural language processing. Whilst this led to considerable advancements, in the main it employed traditional scientific methodologies. However, the challenges and opportunities presented by working with the extremely large datasets of the period has led to new approaches, which led Anderson (2008) to claim that the ‘scientific method’ is “obsolete”. He argues that as opposed to the deductive approach of hypothesis testing, the new big datasets require an inductive approach where correlations are the key to the process:

> “This is a world where massive amounts of data and applied mathematics replace every other tool that might be brought to bear [...] Who knows why people do what they do? The

> point is they do it, and we can track and measure it with unprecedented fidelity. With enough data, the numbers speak for themselves." (Anderson, 2008)

Undoubtedly analysis of big data may lend itself more to inductive approaches than model building, which invariably seeks to reduce data in the interest of model performance and parsimony (Pidd, 1999). However, the demise of the scientific methodology may be somewhat exaggerated, as argued in the many ripostes to the article (e.g. Dyson *et al*, 2008; Granville, 2013). Obviously, correlation depends on linearity, and also the adage that correlation does not mean causality is an important concern. Whilst correlations in big datasets are a valid and growing approach to analysis in the period, in many scenarios a deeper analysis will be more fruitful and appropriate to the problem. It is of course also worth noting, that a biased large sample will be no more fruitful for the analyst than a smaller, 'traditional sized' sample with similar bias.

Decision making maintains equal prominence in the period, with data visualisation attracting much attention and influence. Secondly, there have been many efforts to provide decision support and automation in 'real-time' (e.g. Davenport and Harris, 2007; Niedermann *et al*, 2011). Critical to this is the availability of technology, and specifically processing power. Another key influence has been the effectiveness of search engines in recommending results for users, and the success of product recommendation agents on websites. As a natural extension of this, businesses have sought to provide real-time recommendations for employees such as upsell opportunities and customer churn identification. Such initiatives aim to provide fast, accurate and useful information, improving the speed and precision of decision making (Panian, 2008).

In summary, the sixth and current period has seen new changes to the dianoetic management paradigm, particularly in the form of new data architectures and analytical techniques in response to the abundance of data available. The availability of both the data, and the tools to complement it, have had significant impact on decision makers, the demands of businesses for new and further reaching forms of analysis, and indeed the central methodology of the paradigm itself.

3.2 The Parallel Histories of the Paradigm

The history presented in this analysis charts the development of dianoetic management from time-and-motion studies and basic calculators a century ago, through to the computerised models of the modern age that are automating millions of decisions every second. These changes are evident not only in the physical evolution of the paradigm, the technologies and mathematical models that are used, but also in attitudes to how businesses should be managed. The use of these methods has extended far beyond the factories of Ford and the battlefields of the world wars, into doctor's surgeries, design studios, sports arenas and beyond.

The most obvious and apparent area of growth has been in computing, data processing and telecommunications, with many modern mobile phones boasting 64,000 times the memory of a typical installation of IBM's ground breaking System/360 of the 1960s. The growing amounts of data now available is matched by the growing ability to store, process and analyse vast quantities at ever increasing speeds. Taylor's original calculations, based on sampling the activities of a handful of workers, are in stark contrast to the data-intensive operations of search engines, where simple queries can involve iterating over billions of data-points. Similar progress has been made towards better understanding the decision making process and the effective communication of information. The various disciplines that act both as components and informants of analytics have individually been developing over this period, as has been widely documented (e.g. Kirby, 2003; Ceruzzi, 1999). However, through considering the development of each discipline simultaneously some important interactions and 'ripple-effects' between them can be captured.

As such this represents the first significant contribution of this analysis. Regarding these histories collectively a clear evolution can be seen, with the paradigm growing in both sophistication and in influence. It has been demonstrated that this evolution goes beyond the 'sum-of-its-parts'; as new innovations resonate between each discipline then new applications and opportunities have been exploited and even greater impact achieved. For example, presented in the form of the analysis above, a clear correlation is shown between the growth of 'soft OR' in the 1970s coinciding with the similar growth in influence of many of the 'softer' decision making disciplines and methods into information systems and computer science (e.g. the popularisation of DSS and human-computer interaction). Similar synergies can be seen between the availability of big data, the popularisation of alternative database systems (e.g. NoSQL), and indeed the quantitative methods that Anderson (2008) argues are changing the scientific methodology.

The ripple effects of innovations and influences offer new insights as to the nature of the relationship between the disciplines involved in the paradigm. Whilst each has clear and significant differences, and its own academic tradition and history, equally they are intertwined within an ecosystem. As with any ecosystem, the tendency is to revert to type and maintain its usual practice (its process of "normal science" in Kuhnian terminology). When this is disrupted through new ideas, innovations, and methodologies, the system will seek to adapt and find a new equilibrium, described as a succession in biological ecosystems. This succession is likely to resemble previous states, however, if the scale of the disruption is significant, it is likely to produce significant changes to the ecosystem.

In discussing analytics, we are discussing the current period of the dianoetic management paradigm, and the ecosystem in its current state of equilibrium after the initial disruption of big data and the other factors that marked the beginning of the analytics period. As such we may choose to define "analytics" as simply the most recent and most evolved moment in the history of the paradigm, and the current state of the ecosystem the underlying disciplines co-inhabit.

However, to view the development of the paradigm purely as a straight-line evolution means that the periodization of this history is either irrelevant or solely a convenience serving to carve up this history into more manageable chunks. This conclusion, however, does not seem to fit the data. Of the periods identified, a clear case can be made that each display unique characteristics; are marked by new ideological, methodological and/or technical innovations; and, moreover, have their own preoccupations and causes. In other words, whilst we conceptualise this research as detailing the development of a single paradigm, to more completely describe this history our conceptualisation must also incorporate the separate periods and their individual characteristics.

This conclusion confirms that periodization is not only the product of theory, but it is also a producer of theory (Green, 1995). In what can be considered as the second significant contribution of the research, this facet allows us to generate a new and more satisfactory theory of the relationships between disciplines such as OR and information systems; periods such as analytics and BI; and of the overall dianoetic management paradigm. Whilst many have sought to develop taxonomies that categorise concepts such as MIS, DSS, BI and analytics into super- and subsets, in this theoretical framework such distinctions are in effect not of hierarchy, orientation or methodology but rather they are of *chronology*. In other words, the question is not what differences there are between each, but what concerns, technologies, practices and environmental contexts are prevalent in their time period. Concerns about distinguishing and defining each is more a preoccupation of vendors and academic communities; as Theodore Levitt infamously observed "people don't want to buy a quarter-inch drill. They want a quarter-inch hole" (Christensen *et al*, 2005, p 74).

Similarly, this gives new perspective on the differences between these periods and the associated *quantitative*, *technological* or *decision making* disciplines shown in figure 14, section 2.3. Analytics effectively represents a snapshot in time of the overall ecosystem within which each discipline co-exists. On the other hand, a discipline such as OR represents both a well-established, independent and resilient area of study and practice, which yet also contributes to the dianoetic management paradigm.

To summarise we consider the history charted in this research to be of the whole (the overall paradigm) and simultaneously its sub-sections (each period). This history runs in parallel with the histories of the related *technological*, *quantitative* and *decision making* disciplines from which it draws. The two contributions discussed thus far afford a greater understanding of what analytics actually is, and how it relates to a discipline such as OR. Accordingly, this answers the first research objective (RO1) of the study. However, the second, ascertaining how the OR research community should react to analytics remains unanswered and will be the subject of the next section of this study.

3.3 Implications for the OR Community

The previous section has offered new insight and perspective through considering the history of OR concurrently with the histories of the many other disciplines involved in the paradigm. This shared history not only informs our understanding of how periods such as analytics and business intelligence, but also can be used to infer new insights into the relationship between OR and analytics, and therefore how the OR community should react to its development.

Whilst the *technological*, *quantitative* and *decision making* disciplines associated with the dianoetic management paradigm interact, and notable ripple-effects have been identified between each, within the paradigm itself this is all the more prevalent. In a reciprocal relationship, new techniques and innovations developed in the concurrent histories of its related discipline are absorbed and incorporated into the paradigm. This, in return, affords greater attention and reach into the wider business community for their parent disciplines. These relationships are demonstrated in figure 16.

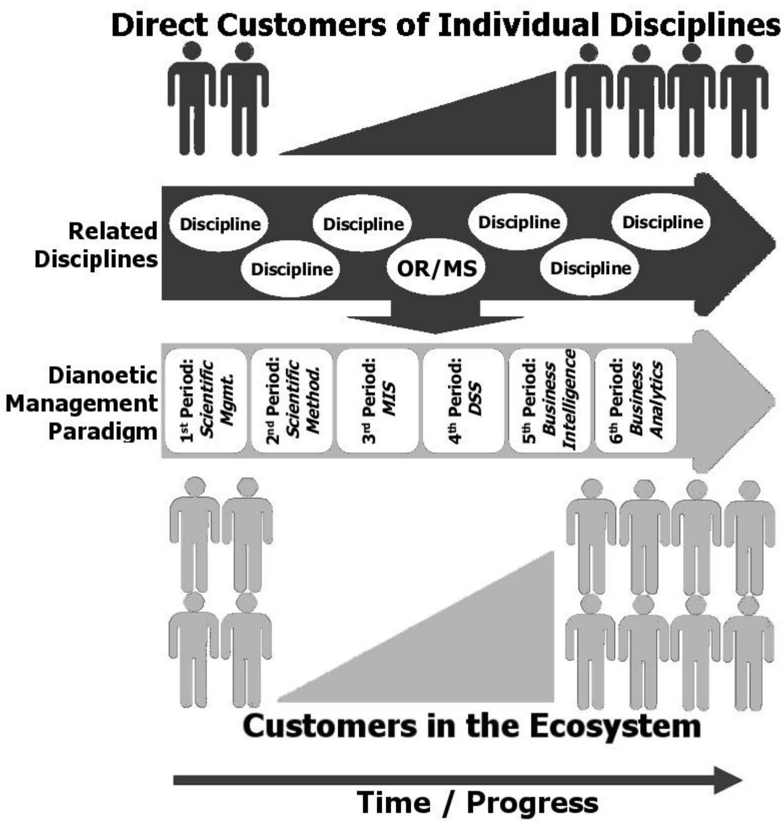


Figure 16 The concurrent histories of the paradigm, periods and related disciplines

The structure of the ecosystem not only gives us insight into how we may distinguish the different disciplines and the periods of the dianoetic management paradigm, but also into the relationship between disciplines such as OR and the others in the ecosystem; relationships which can be both co-operative and competitive. The consequences of this give some indication as to the likely impact of further engagement with analytics may have for the academic OR community. This will be demonstrated by discussing the probable implications of the two extreme positions that the OR community may take in respect to analytics: the *isolationist approach* and the *faddist approach*.

3.3.1 The Isolationist Approach

One option available to the OR community is to distance itself from the overall paradigm. Instead of seeking to engage with each new period or the paradigm as a whole OR can instead focus on best serving its current academic and practitioner communities. An additional benefit of such an approach is to distance the discipline from the uncertainty and hype that is associated with the faster paced developments of the paradigm. However, the trade-off is that the overall reach of the discipline is potentially diminished. As indicated in figure 16, disciplines have their own direct customers who will seek to utilise their methods directly. In the public sector, manufacturing and transportation industries (to name but a few) the OR methodology is relatively well-known and well-used, employing specialist teams and OR consultants alike. However, the number of industries and businesses of which this is true is dwarfed by the number that could benefit from the methods of the discipline. A policy of ignoring both the paradigm and its periods seriously limits the access of OR to the greater numbers of potential customers in the wider ecosystem.

One possible response here would be to argue that the analytics terrain as a whole could be 'captured' by OR, and as such, an isolationist approach is essentially risk-free. Whilst logically correct, the argument seems practically problematic. In the proposed taxonomy of figure 14, OR is cast at the intersection of quantitative methods and decision making. Capturing the entire terrain, by implication, would seemingly also mean capturing the terrain associated with technologies. In other words, a part of OR would have to reposition itself as concerning the computing elements of analytics. Although OR is no stranger to the use of computers, evidence exists that implementations of analytical models into software (including OR-inspired ones) is more often the purvey of IT professionals (Fildes and Ranyard, 1997). Even if this were desired, the relative size of the OR community compared to the computer science community would present a significant challenge here. In other words, even if OR were able to position itself as the principal location for the quantitative and decision making elements, to which it would likely see opposition from communities associated with statistics, machine learning and behavioural science to name but some, to fully capture this space seems improbable.

A similar, more moderate approach may be to seek to separate paradigm and period. The paradigm has drawn upon research, innovations and methodologies from across a spectrum of *technological*, *quantitative* and *decision making* disciplines, of which OR has had a clear, prominent and substantial role. However, does that mean that OR should engage in each of the periods? If the paradigm will continue, and a new period is inevitably around the corner, is it necessary to engage in debate and research into an individual period such as analytics? Irrespective of the appeal such an approach may have, in practice, separating the paradigm from its periods is not so straightforward. The current period is the current incarnation of the paradigm and even if not all of its trends and characteristics resonate entirely with the core concepts of the OR discipline, as concerns of the wider business community they retain relevancy. Whilst each period inevitably gives way to the next, the progresses associated with it continue and are built upon as the paradigm evolves.

Ultimately OR is in competition with many other disciplines for the attention of business users (customers), both now and in the future. Whilst this may seem counter-intuitive to the argument that these disciplines are sharing the same ecosystem, and the reciprocal relationships this entails, the organisms in biological ecosystems compete for natural resources, and with varying degrees of success. To ignore this fact could have highly detrimental results for OR. The devotion the deities of ancient Egypt, Greece and Rome once received did not prevent their decline; a religion without followers soon becomes a footnote in history.

3.3.2 The Faddist Approach

The opposite to the isolationist approach of complete disassociation would be a policy of high engagement and convergence with analytics. This would likely take the form of reinventing the discipline to adopt the new aspects and technologies of the period and renaming many of its societies and publications. Accordingly, the problems associated with the former are reversed; by pushing the connection with analytics, OR could increase its exposure and reach to the considerably greater number of customers in the ecosystem as a whole.

However, there are equally dangers with this approach. Whilst the concerns of the wider ecosystem should therefore have clear relevancy for both the academic and practitioner OR communities, this does not mean that the discipline is, or should be, entirely subsumed by analytics, or that it should seek to entirely reinvent itself. By default, the model necessitates that eventually each period will give way to the next, and the concerns, preoccupations and the terminology will again move on. To have engaged in a complete reinvention can lock the discipline with a moment in time likely to soon be seen as dated and detached from future periods and their principal concerns.

3.3.3 Towards a Balanced Response to Analytics

Both approaches have clear benefits; however, each too carry risks or reduce the potential value business interest in analytics may generate. As with many such situations, the answer probably lies somewhere in the middle. It is important for the OR community to engage with both the paradigm as a whole, and also the current period of analytics, in order to maximise its reach and ensure its relevancy to businesses, practitioners, academics and students. However, it is also important for OR to maintain its distinctiveness and unique selling points so to enjoy longevity and the continued support of its direct customers. Consequentially, a balanced approach is recommended that can both highlight the many qualities and successes of the discipline, as well as engaging with the new concerns of analytics and the wider ecosystem.

This recommendation represents the third contribution of this analysis; an appreciation of the reciprocal relationship between OR and the paradigm re-enforces the need to promote interdisciplinary research and training to the OR community, and to seek to encourage new debate and engagement across the paradigm's business users. This insight would go beyond the concerns of the current period (analytics) into whatever direction the paradigm next goes. Such an approach, however, needs to be enacted across the breadth of the OR community and therefore the lack of academic research is a concern that needs to be addressed. This chapter will conclude by suggesting some specific research themes, thus answering the second objective (RO2): what research directions can be suggested that may unite OR and analytics.

3.4 A Research Agenda for OR in the Analytics Age

As discussed in the previous section, research into analytics should seek to both incorporate the unique aspects of the OR discipline, as well as the innovations, concerns and characteristics of the analytics period. To this end, and to answer the second research objective (RO2), developing an agenda for OR research, five areas of innovation and key developments associated with analytics are suggested as the starting point for future OR research. These areas, by no means comprises an exhaustive list, are: big data, new data architectures, unstructured data, real-time analytics, and data visualisation.

3.4.1 Leaveraging Big Data Volumes

One of the most noted aspects of the analytics age has been the growth in data volume, and in the size of datasets. The latter represented a significant challenge for both the technologies, discussed in the next section, and also the quantitative methods used. One such implication surrounds the use of statistical significance in very large datasets. Whilst a pressing concern traditionally has been collecting enough data to find significant effects, in very large datasets the opposite can be of issue: almost every relationship can be measured as significant at the 5% level. Further research and debate should be encouraged in the wider quantitative community as to what methods can be used for hypothesis testing and model validation in such datasets.

Secondly, and more specifically, the use of big data has significant implications for many of the typical models used in OR practice. Traditionally in such models, simplicity has been advocated (e.g. Ward, 1989), which is not necessarily concordant with using the vast, varied and complex datasets becoming available in the analytics period. To some this may present something of a *Catch-22*: either abandon certain key principals of OR modelling or ignore the potential benefits that big data may bring. However, some practitioner examples are emerging of the use of optimisation techniques in big data (e.g. FICO, 2013). Future research of this kind, or into the limitations and applications of optimisation and other OR techniques to such datasets should be strongly encouraged.

3.4.2 Utilising New Data Architectures

Often synonymous with the subject of big data are the new types of databases, techniques and architectures popularised in the period such as NoSQL, Massively Parallel Processing (MPP) and Hadoop. Whilst in the main such systems are at the more technological end of the spectrum than usually inhabited by OR, that is not to say they are without relevance. As these systems grow in usage in the wider community, or even become de facto, so too does the need to demonstrate how OR applications can be aligned with this architecture. This is not just an ‘emperor’s new clothes’ situation whereby OR can just be transported to new software and new systems, in distributed architecture there is often a need (or opportunity) to rewrite and redesign algorithms. Running algorithms in parallelised computer clusters, where seperate machines are processing separate data simultaneously, and typically not collating results until the end of each iteration, presents opportunity to “scale-up” OR methods to process larger datasets, but requires changes to be made to the procedures. As example, Zhao *et al.* (2009) present a version of the popular K-means clustering algorithm adapted to operate in these environments.

Whilst examples of data mining and machine learning algorithms applied within distributed systems are numerous (e.g. Zaki and Ho, 2000), limited academic literature on the application of

OR methods within these new architectures was found (one example is Taylor, 2018). Further case-studies and reports of experimentation which explore these opportunities are recommended. Such studies can inform the community about these tools, as well as demonstrate their potential benefits and growing prominence.

3.4.3 Incorporating Unstructured Data

As discussed, data in the analytics period has not only been characterised by its unprecedented scale, but also its variety. In particular, this is due to the proliferation of online user-generated content (e.g. blogs, online customer reviews and "tweets"), used for a wide range of tasks such as customer research, epidemiology, security, and risk-analysis. The overarching value inherent in this data lies in the fact that much of it provides highly immediate and uncensored access to the activities, views and interactions of ordinary people. The implications of such access are significant in understanding how social systems work, how information passes through networks and communities, and to predict future events significantly earlier than with traditional data types.

Data of this kind could clearly add additional value in a variety of OR models including simulation, systems dynamics, supply chain management, logistics, and forecasting. As such a variety of research directions in this area should be encouraged: how such data is pre-processed (again dimension-reduction is likely to be necessary due to the sparsity of text and multimedia datasets); how such data can be used effectively in OR models; and case studies demonstrating and/or promoting the use of such data in OR applications.

3.4.4 Streaming Data and Real-time Analytics

An additional consequence of the explosion of online data is that many valuable sources of data are now available online, via application programming interfaces (APIs) or file transfer protocol (FTP) from external websites. This, in combination with ever increasing computer processing power, has significant implications for modelling as it effectively can allow some data collection and processing to occur in close to real-time and 'streams' of data to flow into models autonomously. Meanwhile, real-time applications of OR are relatively prominent in the literature. Examples can be found in various areas including:

- Optimisation (Seguin *et al*, 1997; Diehl *et al*, 2002; Powell *et al*, 2002)
- Simulation (Davis and Jones, 1988; Bruzzone and Giribone, 1998; Better *et al*, 2007)
- Logistics & Scheduling (Seguin *et al*, 1997; Giaglis *et al*, 2004; Durbin and Hoffman, 2008)
- Stochastic Modelling (Davis and Jones, 1988; Sand and Engell, 2004).

Clearly this demonstrates that such research is indeed being generated, and has been for nearly thirty years. Further research may seek to promote this area and bring it to the attention of the wider community, through case studies and/or literature review(s).

3.4.5 Visualising Data

Data visualisation has become one of the main 'buzzwords' of the analytics age, but, as the valuation of Tableau (one of the main software vendors) at $2billion dollars just two days after its initial public offering on the stock market (Cook, 2013) indicates, there has been more to this than just hype. Visualisation is again not necessarily new to the period (Friendly, 2002), but is becoming an area of significant growth partly due to the ability to display visuals on interactive internet browsers, allowing increased distribution and increased power. The potential of these techniques and technologies as tools for effective communication, to increase the impact of analytical findings, and even as stand-alone analysis tools, has been widely acknowledged in the ecosystem at large.

OR, as a discipline closely focused on decision making, can see real benefits from visualisation such as improving the ease of model validation and increased buy-in from stakeholders. Methods such as simulation have long utilised graphical displays for these purposes (Hurrion, 1976) and whilst there has been research into their design and use (e.g. Belton and Elder, 1994), further studies into best practices, in particular with respect to recent visualisation work should be encouraged. Similarly, research into the use of visualisation techniques across the breadth of OR methods, both in theory and practice, may again offer new opportunities for the discipline, and increase awareness amongst OR professionals as well as the wider community.

3.4.6 OR and the Wider Ecosystem

Although not directly linked to OR and analytics, the implication of positioning OR as a constituent member of a wider ecosystem that includes many related disciplines sharing similar goals and concerns equally suggests future work. The main implication of this representation is that, from a business perspective at least, it is in combination that the disciplines can have greater impact and influence. Consequently, a major recommendation would be to encourage future collaborative research between these disciplines, research which could be mutually beneficial for the wider ecosystem, and the prominence, effectiveness and impact of the OR methodology.

One opportunity would be to expand the work started here into a more comprehensive history of the overall ecosystem, particularly in expanding the scope beyond the 100 years explored in this study. Secondly, studies into how the disruptions and ripple effects spread through the ecosystem and how new successions are reached, may shed further light on this phenomenon, as well as inform the disciplines on how to better manage and react to innovations emerging from related disciplines. Thirdly, studies could also focus on the actual process of academic collaboration between these disciplines, with the purpose of identifying barriers and critical success factors, and developing best practice guidelines. Through work such as this, and indeed other opportunities may be identified, a greater understanding of the paradigm as a whole can be reached, an understanding that can help shape the future of the ecosystem rather than simply exploiting the current opportunities it offers.

3.5 Summary

This chapter has reviewed the existing literature concerning analytics, OR, and their relationship; debated some of the reasons why such research has been so limited; and also, some of the broader issues, innovations and implications across a spectrum of disciplines which co-inhabit the same ecosystem. This history has been presented as the paradigm of dianoetic management, defined as the use of *technological*, *quantitative methods*, and *decision making* techniques in order to make business decisions based on data and analyses rather than solely on intuition. The history of this paradigm has been presented as a series of periods, each of which have unique characteristics, whilst simultaneously being part of an overall evolution. Through this analysis an understanding of how analytics relates to related fields and discipline, including OR (thus addressing RO1). Using the themes that are particularly prevalent in the analytics period, examples of possible research directions for the OR community have also been presented (addressing RO2).

Above all the analysis demonstrates that OR does not exist entirely in isolation; the community must embrace and engage with the wider concerns of the ecosystem and paradigm or risk declining into obscurity. With other academic and practitioner communities engaging with analytics and increasing research in these areas, OR is in danger of being left behind. Whilst arguments may be made that such research directions risk diluting the OR 'brand', the original conception of the discipline was to use the most relevant methods available to solve business problems, a tradition such research falls firmly within. However, of course OR research represents just one of the concerns of this study. Another, and perhaps greater concern, is the degree to which teaching programs meet the needs of analytics employers. This issue will begin to be tackled in the next chapter.

4 Not in my Job Description: A Topic Model Analysis of Analytics Job Adverts

Critical to our primary goal of determining the efficacy of university training for analytics professionals, is to determine the requirements of such roles. Accordingly, this not only represents the third objective of this research, but also the focus of two of the research methods employed. In keeping with the pragmatic stance employed in this research, the area will be investigated both qualitatively and quantitatively. The former of these will be through interviews with analytics and OR employers, and is described in chapter six. However, initially we begin this process through a quantitative analysis of analytics job adverts, in comparison with those of related fields (OR included). The methodology of this analysis has been described in section 2.4, and, in short, utilises topic modelling and correlation analysis to contrast analytics jobs with related fields.

Accordingly, the rest of the chapter is arranged as follows. We begin by discussing the data source and pre-processing steps. Secondly, we evaluate the core topics (themes) prevalent in analytics job adverts. Thirdly, a correlation analysis is performed to compare analytics jobs with those of other fields, based on how the different topics rank within the adverts (proportions). Finally, the analysis will focus specifically on the differences observed between analytics and OR adverts, and what this may imply about the relationship between the two. These activities will, in part, address RO3.

4.1 Data Selection, Extraction and Pre-Processing

The first concern is the source for the data for use. In choosing to use the automated approaches of text analytics, it becomes necessary to seek a large volume of data, on the basis that machine learning algorithms, on the whole, tend to improve in predication accuracy as the value of n increases. For this reason, it becomes preferable to use online job boards. Although some degree of content control is lost through this decision (in comparison to selecting job adverts from specialised publications), this is counterbalanced by far greater availability and scale. Whilst there are many such job boards, one of the most popular is that hosted in the popular professional social network www.linkedin.com (which also has the advantage of providing an application programming interface (API) to their job search from which adverts can be extracted).

In order to select relevant job adverts, keywords can be passed to the API (in the same way that a user would specify search terms on the main website). An obvious first keyword is "analytics", using which the first category of job advert was selected. (As a note, the keywords "data science" and "data scientist" were also tested but ultimately very few results were returned that did not also match to "analytics"). Thereafter, we also searched for the term "operational research"; also adding the common US spelling of "operations research" and "management science".

As alluded to in the earlier discussion, however, it is our belief that to construct a meaningful comparison between OR and analytics jobs, we also need to compare these similarities to the similarities that analytics jobs have with other related disciplines. Without doing so there is effectively no yardstick against which we can judge the comparison. Whilst there are a range of reference points that could be used to determine which other disciplines to include, for the purpose of this research we select from the taxonomy of related disciplines shown in figure 14.

As previously stated, we already extract on the terms related to analytics and OR, and in the interests of creating a manageable dataset size and extraction workflow, we limit our remaining selections to one discipline (selected on the basis of the quantity of job adverts available) from each of the remaining sections of the Venn diagram. Whilst this obviously limits the depth of our comparisons, and indeed the list of disciplines is not intended to be exhaustive any way, it does enough to provide a rational comparison point in consideration that our primarily focus is on similarities between OR and analytics.

Also worth noting is that the labels of these adverts are essentially self-selected, they are the keywords that have been included by the job poster. As such, we have no 'quality control' in respect to the appropriateness of these keywords (or otherwise). Inevitably this will introduce noise, however, it is expected that this will be countered by the scale of the analysis; that by collecting a volume of job adverts the 'noise' will be less pronounced. Additionally there were some circumstances where job adverts matched for more than one keywords in our search (for instance, matching both OR and analytics). In these instances the advert was removed from the analysis, as the cost would be complicating any comparison between disciplines. The disciplines (search terms) and the respective quantities of data extracted, broadly representative of the volumes of data available at the time of searching, are shown in table 19.

Table 18 Disciplines selected and advert quantities extracted

Category	*n*
Analytics	12,000
OR	2,745
Statistics	3,593
Machine Learning	4,046
Computer Science	3,888
Information Systems	11,057
Psychology	1,650
TOTAL	**38,979**

Having selected and extracted the job adverts of interest, the next task is to clean and transform the data to make it more amenable for further analyses. This process consisted of the following steps, which are broadly typical of text analytics processes of this kind (e.g. Blei *et al*, 2003).

1. Duplicate removal: removal of adverts that share the same job ID (an internal primary key used in the LinkedIn API) and/or that show a very high level of similarity with others in the dataset;
2. Removal of non-English listings;
3. Removal of punctuation, whitespace, numbers and HTML/XML code;
4. Removal of words less than 3 characters in length;
5. Removal of 'stopwords': removal of words that are very frequent in the English language (such as pronouns, prepositions and conjunctions) which offer little discriminatory power between documents;
6. Removal of words that occur only once in the corpus;
7. Stemming: the transformation of words to their shortest stem, such that the words "operational" and "operations" are reduced to "oper"; and "optimisation" and "optimization" reduce to "optim". Whilst this clearly can reduce the interpretability of the terms, it allows for different tenses, pluralisation's and alternative spellings to be considered as the same word, and greatly reduces computation complexity.

Having performed these steps, the data was ready for the analyses, the details of which are discussed in the remainder of this section.

4.2 Model Build

The instrument used for the analysis is a topic model (based on Online LDA, described in section 2.4.2) of 'analytics' job adverts and related fields. The purpose of the analysis is essentially data reduction, in a style akin to principal component analysis. The topic model finds relevant groupings of words in the documents (adverts) which allow researchers to infer subject matters the documents discuss. Through understanding these subject matters, an understanding of the requirements presented in the job adverts, and to make comparison between those of analytics adverts and its related disciplines, including OR. Statistical comparison of the relative frequency of topics in these disciplines is used to better understand the relationships between them and analytics, and will in part provide insights that may help address RO3 (determining the skills requirements of analytics job roles and the extent to which OR professionals may meet them).

Online LDA has been shown to be effective with just one pass over large enough datasets, but, for greater precision, 10 full passes were performed with chunk sizes of 2,000 documents (with each chunk a random combination of job adverts from each discipline).

The final consideration is the value of *K*, the total number of topics. Much like with related approaches such as factor analysis or *K*-means clustering, *K* must be determined *a priori*. There are variety of approaches available to estimate the 'optimal' number of topics to use (e.g. Minmno *et al,* 2011; Lee and Mimno, 2017). However, in practice optimality can be something of a misnomer. Although the standard model is not hierarchical in the sense that nested topics are not in-built (see Griffiths *et al*, 2014 for an extended model that offers such functionality), in practice there is a tendency to emulate this effect due to smaller topics being incorporated into larger ones when *K* is smaller. As a simplistic example, in a topic model with a high value of *K* there may be a large topic associated with "operational research" and another, smaller topic associated with "optimisation". In a model with a smaller value of *K* the topic of "optimisation" may not be 'found', as only the largest topics would be retained. Although not a mathematical certainty, it would be likely that the content associated with this now non-existent topic ("optimisation") would be instead associated with the larger "operational research" topic.

In other words, the "best" value of *K* can often be closely dependent on the level of granularity sought by the researcher. Therefore, several values of *K* were experimented with, seeking to find an appropriate balance between the level of depth required to distinguish the topics of interest, yet general enough to coherently describe and analyse the results. Through such experimentation, a value of 100 topics was settled on for the final model. Obviously we make no claim to this being the best possible value, but more one that is appropriate to our specific task. In a model with a larger value of *K*, we would find the same topics, albeit alongside other smaller topics.

4.3 Key Topics in Analytics Adverts

After building the topic model, the steps for which are described in section 2.4.2, the key output is the posterior per-topic word probabilities; the most likely words for each topic found. The first task was to visually interpret these for each of the 100 topics to determine its subject matter. In doing so, many can be excluded as irrelevant to the objective of this research. In total 43 of the original topics fell into this category, meaning that 57 topics were retained for further analysis (the full list of topics is included as appendix item D). These rejected topics mainly concern subjects related to the recruitment process itself or to human resources. Whilst such topics are indeed a relevant part of job adverts, they are clearly not particularly helpful to the task in hand. To illustrate this point, figure 17 shows two such topics and the labels given (manually), built as a word cloud using www.wordle.net. To improve readability, the words in these clouds (and the others that follow) have been converted from stems to the shortest, logical version of the words.

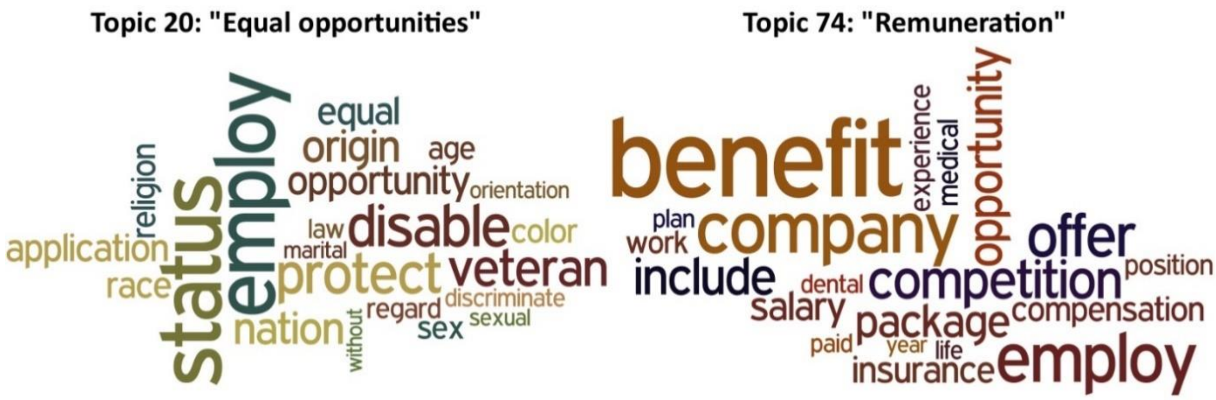


Figure 17 Word clouds of topics 20 and 74

Of course, identifying the subject of the topics is not purely for the purposes of excluding those that are not useful to our goals; these label assignments are critical in all further analysis and, as such, is a non-trivial task, and an integral part of topic modelling. Whilst many topics can be interpreted from the topic clouds alone, for problem cases the posterior matrix can be inspected to find the job adverts that most highly load on the topic for further qualitative analysis. Even after removing the irrelevant topics there remains too many to present here. However, again for illustrative purposes, figure 18 shows four examples of retained topics.

Topic 4, labelled as “analysis (quantitative)”, has, unsurprisingly, quantitative aspects, such as the inclusion of terms such as “statistics” and “data”. This is not to the same extent topic 59, labelled “modelling” and including terms such as “mine”, “predict” and “mathematics”, so we effectively position this as the more ‘descriptive’ end of analytics, further indicated by the prominence of terms such as “report”, the second most likely of all. Topic 9, “analysis (business)”, on the other hand, features terms such as “require” (possibly also linked to “requirements”), “design” and "document" which suggest key steps in the task of business analysis. Such label assignments demonstrate one other important aspect in this task, recognising that a prominent term in both, "analysis", can have a different meaning based on the associations we draw from the other terms; in the case of topic 4 we assume this to mostly refer to quantitative analyses, whereas in 9 this seemingly refers to qualitative (business) analysis.

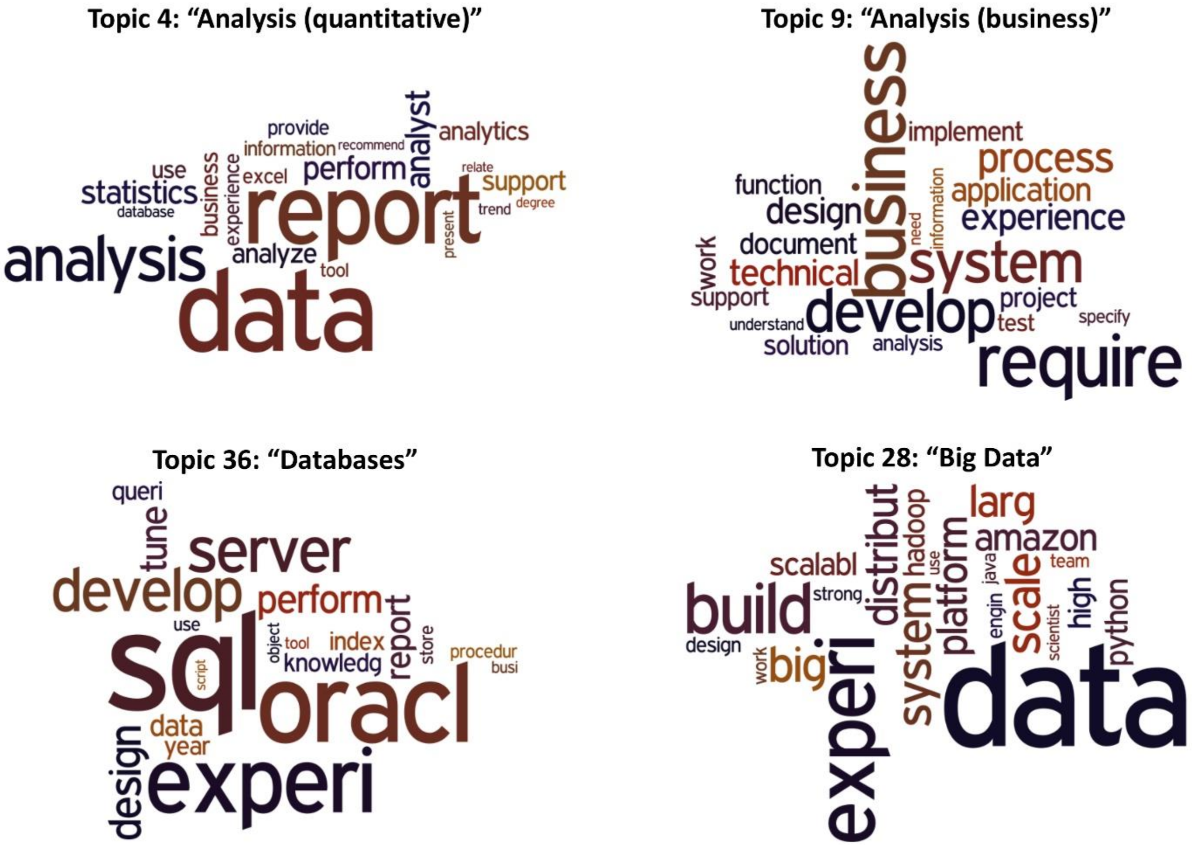


Figure 18 Word clouds of topics 4, 9, 36 and 28

Similarly, the word clouds of topics 36 and 28 (figure 18) also show close similarity, but at the same time can be distinguished into two relatively defined categories. Topic 36 features "SQL" and "Oracle" as its two most prominent words, as well as terms such as "queri[es]", "index" and "procedur[e]", suggesting this topic is related to databases. However, this may be moreover relational databases, as topic 28 incorporates terms such as "distribut[e]" and "Hadoop", associated with NoSQL data storage, along with "java" and "python", both languages used in such architectures often in place of traditional SQL (although a variety of SQL-like languages now exist in such environments). Accordingly, the topic is labelled "Big Data", with acknowledgement that this moreover deals with its management (and is therefore comparable to topic 36). These examples give some illustration of the interpretation process. Whilst obviously subjective, the topic labels are broadly considered to be discrete enough to use in the analysis. (The interested reader may wish to consult Chang *et al.* (2009) which discusses the area of topic interpretation in more detail).

After labelling the topics, and therefore identifying the skills featured in the adverts, our next goal was to identify those which ranks most highly in analytics job adverts. The top 15, based on the topics prominence in the analytics corpus (the averaged θ_i values), is shown in table 20. One immediate observation is that there is indeed a mixture of business orientated skills (with "management (skills)" ranked highest and "communication skills" third); computing skills ("programming" eighth; "software (development)" tenth); and quantitative skills ("analysis (quantitative)" third; "modelling" eleventh). However, beyond these more expected results, there are also many skills we may link with marketing, sales and eCommerce. Whilst these were identified as key domains in the literature, with (arguably) a third of the top 15 associated with these areas, this connection is very prominent.

Table 19 Prevalent analytics skills/domain topics compared to other disciplines

	Analytics	Operational Research	Statistics	Machine Learning	Computer Science	Information Systems	Psychology
Management (skills)	1	=	↓ 2	↓ 6	↓ 6	↓ 2	=
Marketing	2	↓ 4	↓ 5	↓ 8	↓ 9	↓ 15	↓ 5
Communication skills	3	=	↑ 1	↓ 5	↓ 4	=	↑ 2
Analysis (business)	4	↓ 7	↓ 7	↓ 9	↑ 3	↑ 1	↓ 7
Analysis (quantitative)	5	=	↑ 3	↓ 10	↓ 12	↓ 10	↓ 9
Management (teams)	6	=	=	↓ 11	↓ 10	↑ 5	↑ 3
Sales skills	7	↓ 13	↓ 11	↓ 16	↓ 21	↓ 20	↓ 11
Programming	8	↓ 23	↓ 20	↑ 7	↑ 1	↑ 6	↓ 18
Consulting	9	↓ 10	↓ 15	↓ 15	↓ 17	↓ 11	↑ 8
Software (development)	10	↑ 8	↑ 9	↑ 1	↑ 2	↑ 9	↓ 14
Modelling	11	↑ 2	↑ 4	↑ 2	↓ 22	↓ 21	↑ 10
Marketing campaigns	12	↓ 22	↓ 17	↓ 24	↓ 37	↓ 36	↓ 15
Solutions & architecture	13	↓ 24	↓ 32	=	↑ 7	↑ 8	↓ 28
Project management	14	↑ 9	↑ 13	=	↑ 8	↑ 7	↑ 12
Ecommerce	15	↓ 34	↓ 25	↓ 23	↓ 31	↓ 40	↓ 22

There may be multiple reasons to explain this. Firstly, this may be linked to the growing amounts of data available in the modern era relating to consumers and customers. One of the key sources of big data, is user-generated content via social networking sites, review websites and mobile devices. If indeed we accept that analytics is, in part at least, linked to big data, it naturally follows that a significant proportion of analytics roles that utilise big data will be for customer-orientated domains. Secondly this may be representative of the 'verticalization' seen to be a major driver of analytics.

As demonstrated, purely evaluating the rankings of these skills offers benefits. However, as previously discussed, a clearer picture of the analytics skill-set may more easily be reached by comparing it to related fields. Accordingly, table 20 additionally displays the rank of the skills within the other disciplines in the analysis, indicating whether they are ranked the same (denoted by "="), lower (shown as "↓" followed by the ranking of the topic) or higher ("↑").

In comparison, the rankings for OR in respect to the top six topics are reasonably similar. Four are listed in the same spot with only "marketing" and "analysis (business)" listed lower. There are, however, several skills that are ranked considerably higher in analytics than in OR. Programming ranks at eighth in analytics, echoing some of the earlier discussion on data scientists, whilst in OR it ranks at twenty-third. "Solutions and architecture", which relates to the larger scale information architecture of projects and business systems, ranks at tenth for analytics, but twenty-fourth for OR. OR is also considerably lower for the domains of "marketing campaigns" and "ecommerce". In the opposite direction, the most notable difference is that OR has a far higher ranking for "modelling" (ranked second in comparison to analytics' eleventh). If, as suggested in the earlier discussion, "modelling" can be regarded as the more technically advanced in comparison to "analysis (quantitative)", analytics roles may be interpreted as being less likely to use advanced mathematical approaches than OR, machine learning and statistics roles (all of which rank "modelling" notably higher).

The results for statistics invite many of the same comments as for OR. Again, there are notably lower ranks for "programming", "solutions & architecture" and the "ecommerce" domain. "Modelling" is also listed higher (fourth versus eleventh). Interestingly, compared with analytics and OR, there is less focus on management and consultancy skills, whilst "communication skills" ranks highest. The argument of Laney and Kart (2012), that analytics professionals are more likely to work in teams than statisticians, receives mixed support. There is strong evidence for a need for skills related to teamwork in analytics jobs, however, "communication skills" ranking may suggest that statistics roles are not solely within the confines of the 'boys in the back room'. Machine learning shows significant differences too. Perhaps unsurprisingly as a branch of computer science, skills such as "software development" (first) and "programming" (seventh) rank highly, but again there is a greater prominence on "modelling" (second) than with analytics. In general, it is the 'people' skills (e.g. "management (skills)", "communication skills" and "sales processes") which rank lower here than with analytics.

Of the remaining three, a greater amount of variation is visible. Also, much of the results would match a priori theory. Computer science ranks higher on aspects such as "programming" and "solutions & architecture"; lower again on 'people' skills. Information systems ranks higher on many of the same aspects as with computer science, but also maintains high rankings for 'people' skills and business skills. The discipline though is markedly lower on quantitative skills. Psychology, which represents the lowest correlation, ranks management and communication skills higher, but is lower on almost all of the more technical skills.

It is, of course, necessary to retain some criticality towards these findings. As discussed in section 2.4.2, results based upon rankings are inevitably weaker than those based on continuous data (the raw percentages), particularly as we have no measure of the difference between rankings. The gap between the 1st and 2nd skills in a set of adverts, for instance, may be fractionally small, whilst the gap between 2nd and 3rd could be comparatively large (also limiting the potential for treating the data as interval rather than ranked). This is certainly problematic, but it has been decided to be "less worse" than mischaracterising and misusing the data as either continuous or interval. However, the results discussed above, and in the remainder of the chapter need to be considered as weak indicators, rather than statements of facts. These can be validated (or otherwise) by the evidence generated from the other analysis methods presented in the thesis (particularly those of chapter six which looks at similar issues but based on qualitative data collected in interviews).

With these caveats in mind, these comparisons can still give some picture of the nature of analytics job roles. The suggestion is that the role of the analytics professional is typically business and/or client-facing (with the emphasis on business and communication skills), more so than in typical machine learning, computer science and information systems roles. However, there is evidence of a greater emphasis on computing skills such as "programming" and "solutions & architecture" than quantitative disciplines such as OR and statistics. Indeed, there is an overall impression of analytics representing a composite of many of these fields, with a more even spread of skills between the three different areas (technology, quantitative and business) than is suggested in the other disciplines.

4.3.1 Programming Languages

An additional consideration that emerged from the literature review was programming languages, further justified by the relatively high ranking for "programming" skills in analytics jobs (8th overall). The distribution of languages is not easily captured in the topic results in this case, so the untransformed dataset was searched and the frequency counts calculated for each of the top 100 languages in the Tiobe index (http://www.tiobe.com/tiobe_index), ranked by perceived popularity. The prominence of each language by discipline is shown in table 21.

Again, some caution needs to be taken when interpreting the table, due to the issues regarding ranked data. The most frequently mentioned in analytics job adverts, as it is in almost all cases, is SQL (OR and machine learning both list it second). Java is the second most common in analytics, in keeping with the technology-orientated disciplines of machine learning, computer science and information systems. It is listed slightly lower (4th) in OR and statistics roles. On the other hand, SAS shows strongly in analytics (3rd), in common with OR, statistics and psychology, but less so in the technology-orientated fields. SAS' principal competitor (arguably), SPSS, features in the top 10 for OR, statistics and Psychology, but not for analytics or the other two disciplines.

Table 20 Most frequently requested programming languages by discipline

	Analytics	Operational Research	Statistics	Machine Learning	Computer Science	Information Systems	Psychology
1st	SQL	SAS	SQL	Java	SQL	SQL	SQL
2nd	Java	SQL	SAS	SQL	Java	Java	SAS
3rd	SAS	R	R	Python	.Net	.Net	SPSS
4th	JavaScript	Java	Java	R	JavaScript	SAS	HTML
5th	HTML	SPSS	C	C++	C++	C#	R
6th	R	Matlab	SPSS	SAS	R	R	CSS
7th	.Net	Python	Python	JavaScript	C#	JavaScript	Python
8th	CSS	.Net	VBA	Ruby	Python	C	C
9th	C	C++	Matlab	C#	C	HTML	Icon
10th	C++	C	C++	C	HTML	PL/SQL	Java

In comparison to OR and statistics, the web languages of JavaScript, HTML and CSS are prominent in analytics (although .Net features similarly for both OR and analytics), as they are in machine learning and computer science. Perhaps the most surprising result, considering the earlier literature, is that R and Python are relatively infrequent in analytics jobs (with Python not even making the top 10). The juxtaposition of this finding with the discussion in the literature around the skills required of data scientists, may suggest that there are differences between these roles and that of the analytics professional (although data scientist jobs extracted in the API search also matched for the "analytics" keyword). Although beyond the scope of this study, one possible suggestion is that we consider "data scientist" as a subset of the analytics profession.

4.4 Correlation Analysis

Thus far the analysis has found insights in comparing analytics to the other disciplines in a purely qualitative way. However, as ranked data, Spearman's correlation can provide a measure of the overall similarity between disciplines. In effect, this provides a metric that represents the degree of similarity between analytics job adverts and adverts associated with each of the other disciplines; and, therefore, a proxy measure of how similar jobs in each field are with analytics roles. To initiate this, we rank each retained topic by discipline, and compare each ranking with the ranking for analytics. This is firstly presented across the full dataset of relevant topics (i.e. the 100 topics less the 43 removed as note being related to skills). The results are shown in table 22.

Table 21 Correlation coefficients (to analytics) by discipline

	ρ	p-value
Statistics	0.8964	0.000
Machine learning	0.8611	0.000
OR	0.8489	0.000
Computer science	0.8008	0.000
Information systems	0.7876	0.000
Psychology	0.7378	0.000

Unsurprisingly, and substantiating our *a priori* theory, all the disciplines show positive correlation, and at a relatively high level (the range of possible outputs is -1 to 1). Three disciplines in particular – statistics, machine learning and OR – show a positive correlation of 85% or more (with rounding), of which statistics is shown to be the most similar, with a correlation near 90%. In respect to one of the main goals of this research, a subject discussed further later in this section, OR's ranking as only the third most aligned to analytics is potentially concerning for its community, and conflicting with some of the earlier literature. However, the gap between the top three disciplines is very small and insignificant (less than 2% difference between the "machine learning" and "OR" categories).

To expand further on this initial analysis, the topics can be split into the sub-groups hard skills, soft skills and domains (application areas and industries), to mirror the themes of the earlier literature review (section 1.4.4). Thereafter, the analysis of ranked topics is utilised (via Spearman's correlation coefficient) to compare the relative distribution of topic groups between analytics roles and those of the other disciplines. In most cases separating into these groups is relatively straight-forward, though there are a few that may be contentious, such as "project management" and "visualisation" (both of which were ultimately classed as soft skills). A full list of topics and their classifications is shown as appendix item D. Having done so then correlation analyses can be performed, shown in figure 19.

Hard Skills

	ρ	p-value
Statistics	0.8000	0.000
OR	0.7877	0.000
Machine learning	0.7211	0.001
Information systems	0.7088	0.001
Psychology	0.6789	0.001
Computer science	0.6368	0.003

Soft Skills

	ρ	p-value
Statistics	0.9286	0.001
Machine learning	0.8810	0.004
OR	0.8810	0.004
Information systems	0.8333	0.010
Computer science	0.7381	0.037
Psychology	0.6905	0.058*

* no significance at the 95% confidence level

Domains

	ρ	p-value
Statistics	0.8759	0.000
Machine learning	0.8729	0.000
Computer science	0.7793	0.000
OR	0.7522	0.000
Psychology	0.6734	0.000
Information systems	0.6113	0.000

Figure 19 Correlation coefficients for analytics adverts in comparison to other disciplines (split by hard and soft skills, and domains)

Statistics shows the greatest correlation across all three sub-groups, supporting the result of the overall analysis (table 22). OR is second in both hard and soft skills (the latter jointly with machine learning) but drops down to fourth in respect to domains (and with more substantial a difference). Machine learning shows correlation, second for soft skills and domains, and at third for hard skills. The other three disciplines generally score lowest (though computer science is in third for domains). To further explore this, figure 20 shows the average topic proportions for each of these sub-groups. Though, as discussed in the methodology, some caution must be taken in using topic proportions in this way, this does further illustrate some of the above points.

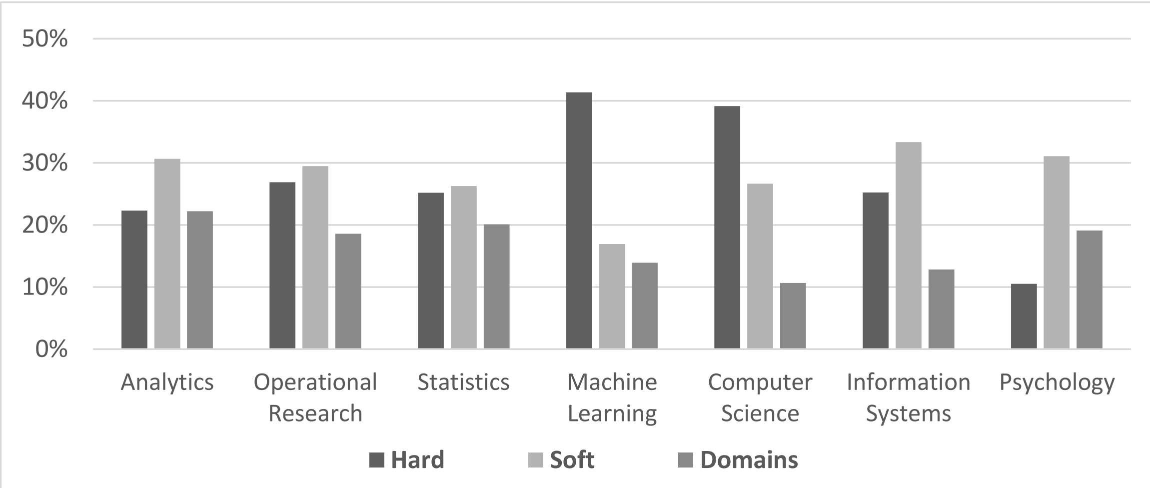


Figure 20 Comparison of average topic proportions by discipline

In the main, analytics, OR and statistics all show a relatively similar composition, although the latter two all have a greater proportion aligned to hard skills. Machine learning and computer science seemingly both have a notably higher emphasis on hard skills, whereas psychology is the opposite. Information systems, in comparison to analytics, have a notably lower emphasis on domain skills, as with machine learning and computer science, but unlike these disciplines there is a greater emphasis on softer skills, as would probably have been forecasted.

To summarise these initial analyses, analytics roles are seemingly most closely aligned to those of statistics, in terms of hard skills, soft skills and domains, followed by machine learning and OR. In the case of machine learning, the types of hard skills required are where the largest differences are seen (a correlation of 72% compared to an overall correlation of 86%). For OR, it is domains where the correlation is weakest, 75% correlation compared to an overall of 85%.

In respect to the balance between of the Venn diagram of figure 14 (section 2.3), *technologies*, *quantitative methods* and *decision making* skills, analytics roles seemingly have the most even balance between the three. However, there is some inference that these may not be in the same depth as some of the other discipline groups. There are two possible interpretations of this insight. One is that analytics roles require more versatility than with some other disciplines, with candidates required to show aptitude in a range of areas, albeit, maybe not to the same concentration (i.e. breadth over depth). The second interpretation would be that analytics roles are more varied in their requirements, with some analytics specialists required to specialise in different areas dependent on the company and the role. In all probability, there will be some degree of both.

The extra breadth is validated in the literature, as shown in the discussion of section 1.4. However, at the same time, there may be some degree of employers, candidates and specialists seeking to "cash in" (whether unintentionally or more cynically) on the growing status of analytics and the data scientist title, with the potential that some of the roles are effectively no different from those in other disciplines in anything other than their title.

4.5 Comparing Analytics and OR Job Roles

As discussed in section 4.2, the analyses point to a relatively strong correlation between analytics and OR roles, in particular in terms of the distribution between hard, soft and domain skills, as well as the correlations between the two, particularly around hard and soft skills. There is therefore, enough evidence to suggest that analytics and OR roles are closely aligned, but by also referring to other disciplines we can view analytics as something of a composite.

- Analytics and statistics share communality across skills and domains;
- Analytics aligns with machine learning roles, particularly in its emphasis on programming, web and software and architectures;
- Analytics broadly overlaps with computer science, information systems and, to some extent, psychology (to varying degrees);
- Analytics and OR share a similar balance between hard and soft skills, and potentially relate in how they integrate and impact with business process.

Overall, the perspective of analytics the research suggests is that of an inherently interdisciplinary practice, probably even more so than a discipline such as OR. Analytics roles can incorporate technology skills (from programming to IT architecture), quantitative analyses and modelling, and suggest a considerable business orientation (with management skills ranked as highest of all).

Empirically evaluating the similarities between OR and analytics roles, is not immediately obvious from the results of the previous section. OR and analytics show a correlation coefficient of around 0.85, which is clearly relatively high. Statistics reports the highest correlation (marginally higher than OR), not only in the overall rankings, but also in the individual sub-categories of skills and domains. In other words, whilst we see a substantial overlap between OR and analytics roles, the discipline is not the only show in town. Analytics also shares much with statistics and machine learning roles, albeit potentially of the four the average analytics job may not feature the same technical complexity in terms of mathematical modelling.

Further analysis can help illustrate where some of these differences, between OR and analytics roles, can be found. As per the analysis of the literature, there are sub-facets within this including the balance between soft and hard skills, the different domains which the jobs feature in, and the programming languages most frequently used. The major differences in these regards are shown in tables 23 and 24.

Table 22 Comparison of analytics and OR skills

	Hard Skills	Soft Skills	Domains	Programming
Similarities	Analysis (quantitiative); software development; Big Data	Management (skills); communication skills; consulting	Marketing; financial (control); financial (audit);	SQL; C; C++
OR+	Modelling; machine learning; process monitoring	Project management	Manufacturing & SCM; intelligence & operations; engineering & safety	R; SPSS; Matlab
Analytics+	Programming; solutions & architecture; business intelligence	Analysis (business); sales skills	Marketing campaigns; ecommerce; advertising	Java; JavaScript; HTML

Table 23 Comparison of hard and soft skills, and domains: Analytics and OR jobs

HARD SKILLS		SOFT SKILLS		DOMAINS	
Analytics	**Operational Research**	**Analytics**	**Operational Research**	**Analytics**	**Operational Research**
Analysis (quantitative)	Modelling	Management (skills)	Management (skills)	Marketing	Marketing
Programming	Analysis (quantitative)	Communication skills	Communication skills	Marketing campaigns	Manufacturing & SCM
Software (development)	Software (development)	Analysis (business)	Management (teams)	Ecommerce	Financial (trading)
Modelling	Machine learning	Management (teams)	Analysis (business)	Financial (control)	Financial (control)
Solutions & architecture	Process monitoring	Sales skills	Project management	Financial (trading)	Financial (credit)
Systems management	Systems management	Consulting	Consulting	Customer management	Public sector (governing)
Process monitoring	Big Data	Project management	Sales skills	Public sector (governing)	Customer management
Big Data	Programming	N/A	N/A	Advertising	Marketing campaigns
Product development	Solutions & architecture	N/A	N/A	Publishing	Intelligence & operations
Business intelligence	Research	N/A	N/A	Human resources	Engineering & safety

In respect to hard skills, OR and analytics show similarities in several areas, but also clear differences. OR seemingly has a greater emphasis, as has already been suggested, on more advanced mathematical approaches, whereas analytics has more of a focus on IT skills. Soft skills show a much closer match, with only project management more frequently required in OR roles, and business analysis and sales skills showing more prominence in analytics. However, it is in domains that by far the biggest differences can be seen, as was shown in the considerably lower similarity metric in figure 19. OR has stronger connections to industrial sectors such as manufacturing, supply chain and engineering, as well as military operations and intelligence, and government. Analytics, on the other hand, shows a strong affiliation to marketing, advertising, sales processes and, in particular, the digital world, with ecommerce, social media and digital marketing all ranking considerably higher.

Both disciplines show a strong emphasis on the SQL and SAS programming languages. However, OR features several other statistical and mathematical languages (such as R, SPSS and Matlab) far more prominently, whilst analytics, most notably, shows a more substantial emphasis on web languages. This may be partly explained by analytics' closer association with digital domains, but also potentially the use of these technologies in visualisation (which also scores more highly for analytics than OR). JavaScript, in particular, is widely used to this end, such as the popular D3 library (https://d3js.org/).

In comparison to the literature, validation can be found to some of these ideas. For OR skills, our highest ranked topics, "management (skills)", roughly corresponds to the function most respondents reported they worked with ("senior management") in a survey of OR practitioners in the UK (Carter, 1987). The same survey ranked "marketing" as the fifth most common function, the fourth ranked skill in this analysis. Writing a little over a decade later, a similar survey (Fildes *et al,* 1999) found marketing to be the second most common function. Their research also find results that support some of our findings, such as the role of OR professionals, and skills requirements, in software development and project management (the 8th and 9th ranked skills).

In respect to the literature regarding analytics roles, the results are less clear. "Programming" ranked considerably higher for analytics in comparison to OR (8th compared with 23rd). This would seem in keeping with much of the discussion of section 1.4, which seemed to suggest that programming was becoming increasingly important (for instance, Davenport and Patil (2012) that a "data scientists' most basic, universal skill is the ability to write code"). However, this seems contradictory to the survey comparison between analytics and OR professionals presented in Liberatore and Luo (2013), which found programming to be ranked higher for OR professionals than their analytics counterparts. Similarly these authors found "project management" more important to analytics professionals than OR, directly opposite to our findings (where they were ranked 14th and 9th respectively). The ranking of "communication skills" as third highest for analytics roles (the same as for OR) seems also to recall this literature, for instance Laney and Kart (2012) suggesting "data scientists are expected to work more within teams than statisticians". However, in direct contrast to this, "communication skills" actual ranked higher for "statistics" roles, as the highest ranked of all in the discipline.

In summary, whilst these results will need further evidence to support them before they can be reliably accepted (a task that will be assumed in chapter six), the findings so far suggest several shared skill requirements between OR and analytics, as well as several points of difference. The implications for the OR community are several. Firstly, if we do indeed accept the premise that analytics and data science offers further opportunities for OR professionals, then one key area for potential upskilling is in programming and IT development. Web technologies are seemingly high in demand. However, most notably, it is within domains that the biggest differences can be seen, such as an extra emphasis on domains around marketing and the web for analytics, in comparison to the heavy industries and public sector roles found more regularly in OR roles. This may have implications for practitioners, in terms of the sectors that can be targeted (reminiscent of the earlier literature suggesting the growth of analytics has, in part, been bought about by a growing verticalisation through industries). However, it also has implications for educators and researchers to ensure that their case studies, datasets and examples can replicate this increased diversity of applications.

4.6 Summary

The analyses of this chapter describe the most frequent skills requirements in analytics roles, and provide a range of different approaches to compare analytics with OR and also other disciplines. The argument of the previous chapter, that there is a relatively strong relationship between OR and analytics, is further evidenced in this chapter, although there is equally evidence, indeed stronger evidence in some cases, for analytics' relationship with other disciplines. OR has been shown to be relatively close in terms of the hard skills and soft skills required, but weaker in the domains associated with it, and a lesser emphasis on digital, marketing and web in particular.

These analyses alone are not adequate to fully address RO3, however, they do provide several insights, and help develop the questions for the interviews with analytics and OR employers. These interviews are discussed and analysed in chapter six, but before this the next chapter investigates the current state of analytics and OR education in UK universities.

5 Degrees of Separation: A Typology of Analytics Master's Degrees in UK Universities

Chapter four has provided an in initial analysis of the requirements of analytics and OR employers, one which will be complemented by the further qualitative analysis of chapter six. However, the other core component of the research question relates to the teaching and training of graduates to meet these demands. In recognition that there is not, and probably should not be, a single, one-size-fits-all approach to such endeavours, this chapter will seek to explore the breadth of current provisions through a quantitative analysis of online degree materials and curricula. In particular, we seek to compare the core elements of degrees titled “Analytics” (or similar) with OR degrees and many of the other related disciplines identified. To this end, we build on the discussion of section 1.5, which detailed prior research into the disciplines related to analytics and the core components of such degrees, by utilising bagged classifiers and the weighted module analysis derived for this research (both detailed in section 2.5).

The chapter is organised as follows. We begin by describing the data collection and processing steps and model specification. Secondly, we present the results of the bagged Support Vector Classifier (SVC) for classifying analytics degrees to the most similar degree title of related disciplines. Thirdly, we perform the module analysis of analytics degrees in comparison to the same disciplines. Fourthly, we seek to synthesise these findings to develop a typology of analytics degrees, before the final section concludes the chapter, and considers the implications of these findings.

5.1 Data Extraction, Pre-Processing and Model Build

The overall approach, therefore, was to use a range of SVC classifiers, built with bagging, and use these to predict a class label for course materials associated with master's degrees titled "analytics", "data science" or similar (e.g. "business analytics" or "big data analytics"). In other words, we are seeking to classify analytics-type degrees based on their similarity to traditional academic disciplines, the results of which can indicate the relative prominence of the teaching of these disciplines. The class labels were to be drawn from a range of disciplines related to analytics using the taxonomy of disciplines shown in figure 14 (section 2.3), the same as are used in 2.4 (the job advert analysis). Namely these are OR, statistics, machine learning, computer science, information systems and psychology.

The degree materials were collected by manually checking each UK university's website directly (using the Guardian's University League Table (http://www.theguardian.com/education/ng-interactive/2014/jun/02/university-league-tables-2015-the-complete-list) to determine the institutions to include) and extracting written materials on the relevant masters-level courses. In cases where universities offered more than one degree with a related title then only the one deemed most relevant was used (e.g. "Business Psychology" was deemed more relevant than "Clinical Psychology"). In cases where 'hybrid' degrees were offered, such as an MSc in "Operational Research and Applied Statistics", the result was excluded on the basis this would complicate classification. In doing so, obviously potentially useful data has been excluded, particularly relevant as compared to the job advert data collected, this dataset is considerably smaller. However, if this were included it would by default sit between two classes, and not easily managed. The options would be to include it twice (which would be very problematic for learning class differences) or to divide the materials into two manually (requiring subjective judgement much of which is problematic). In other words, although we lose potentially useful information in excluding these results, the alternatives seem worse. Additionally, we also collected materials from degrees related to analytics, including titles such as "Business Analytics", "Big Data" and "Data Science". In this instance, multiple degrees for the same institution as well as 'hybrid' degrees were retained, as this information was considered useful to our goal, and the results were qualitatively assessed after the analysis.

As discussed above, for the SVC model the course material data was supplemented with job adverts linked to each discipline which is used as a proxy due to the relatively small size of dataset. The job adverts were drawn from the same database as used in the job advert analysis, detailed in section 4.1, and we followed the same cleaning steps given in this section. As SVC models are sensitive to class sizes, and can perform badly when there are significantly more of one class than another, 1,500 job adverts were randomly selected from the keyword searches made for each discipline. The total quantity of all documents extracted is shown in table 14.

Table 24 Quantities of documents collected by discipline and type

Discipline	Code	Job Adverts*	Degree Materials
Computer Science	CS	1,477	69
Information Systems	IS	1,494	40
Machine Learning	ML	1,415	16
Operational Research	OR	1,481	10
Psychology	PS	1,500	37
Statistics	ST	1,479	19
Analytics	AL	---†	43
TOTAL		**8,846**	**234**

* The quantity of job advert data varies due to the removal of duplications

† Analytics job advert data was not extracted as this is not used as part of the model build, and solely as the 'validation' set

An additional issue with the use of both SVCs and bagging methods is the selection of certain free hyperparameters, the choice of which can significantly impact the accuracy of the model. The most important parameters for our SVC model are C, the regularisation penalty of the error term, and the tolerance of the stopping criteria. For the bagging meta-algorithm, our principal concern is the number of separate estimators to build, and the number of samples to draw from the training set. These parameters were optimised using grid search (e.g. LeCun *et al*, 1998), whereby different values are tested on different folds of the data (using cross-validation), and the values which produce the highest averaged accuracy are retained.

Our final concern is data quantity. For this study, we chose to limit the course materials to master's degrees offered at UK universities. This was partially due to reducing the possible variation that may come from national and/or language difference between materials from universities from different countries, and partially due to the time-consuming nature of collecting data of this kind. However, the accuracy of SVC predictions tends to increase as the size of the training dataset increases, and with only 234 course descriptions the size of the data set would be considered relatively small.

Our solution is to seek data that can be used as a 'proxy' for course descriptions; documents that can be classified in the same way (by the disciplines above) and which display comparable characteristics as degree materials in respect to the similarities and differences observed between categories. A likely candidate for such criteria is job adverts related to the respective disciplines (as has been utilised in the job advert analysis). It could be assumed that both OR degree materials and job adverts will include terms such as "optimization" and "simulation" at a far higher frequency than those associated with information systems, and, vice versa, they are less likely to feature terms such as "data warehouse" or "ERP". Whilst, intuitively, job adverts and degree descriptions would read very differently, this should have negligible impact on their usefulness for this task. Essentially the central mechanism of the model is based upon the terms which most distinguish the different categories, not the terms that distinguish the documents. In other words, it is the relative use of terms such as "optimisation" and "ERP" that will distinguish class membership, not the relative use of terms such as "salary" and "lectures".

Of course, the implicit assumption is that the master's degrees linked to each discipline are actually aligned to the jobs they are intended to prepare students for, an assumption that some may well question. Therefore, to validate job descriptions as suitable surrogate for the course materials, the model, trained on job advert data, can be tested against a subset of the job advert data as well as the course materials. If the accuracy of the model does not significantly decrease when using the course data to test the model, the job advert data can be taken to represent an appropriate proxy.

The full process for this research method, therefore, is as follows (and represented in figure 21):

1. Fit the model using a subset of the job advert data associated with each discipline (the training set), each of which represent separate classes;
2. Optimise the hyperparameters using grid search;
3. Use the model to predict the remaining subset of job advert data (the test set) and evaluate its accuracy;
4. Use the model to predict the course materials associated with each 'traditional' discipline (the de facto validation set) and compare its accuracy to that of the test set;
5. Use the model to predict the classes for analytics-type degrees and evaluate the results.

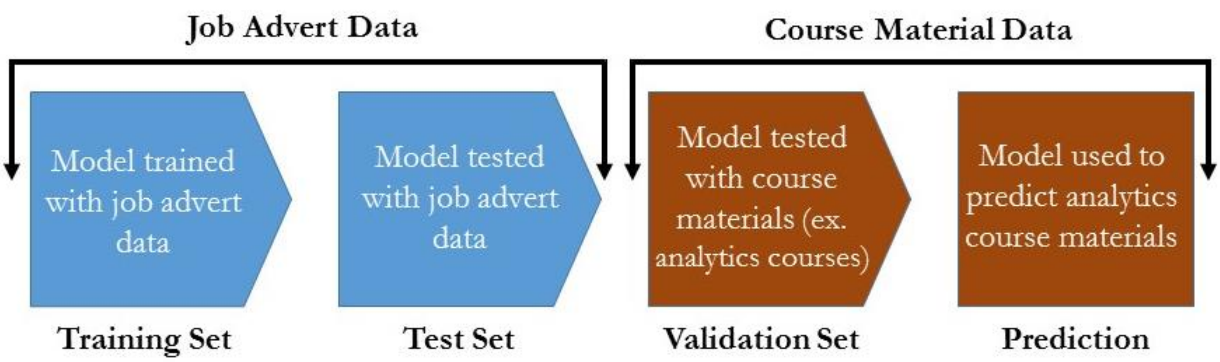


Figure 21 Data used in the SVC model

These procedures were performed in Python, using the scikit-learn (Pedregosa *et al*, 2011) and NumPy (Van der Walt *et al*, 2011) packages and the Python Data Analysis Library (PyData, 2012).

5.2 Classifying Analytics Degrees

As detailed in the introduction, the first task of this chapter is to seek to identify the degrees with which analytics MSc degrees most closely align with. The value of this is to consider different possible 'flavours' of analytics degree, and to in some way quantify the similarity of analytics teaching with that of OR and related disciplines. As before, the pool of degrees to which we draw these comparisons, are the same as used in chapter four (and are based on the taxonomy of figure 14, chapter two). Namely these are OR, statistics (ST), machine learning (ML), computer science (CS), information systems (IS), and psychology (PS).

As per the steps listed in the previous section (5.1), the first stage was to divide the data so to create a training set of job advert data (upon which the model is built); a test set (the remainder of the job advert data) to confirm the model can predict the correct class using the same data source; and finally, a validation set of degree materials (to confirm the model can also adequately predict the correct class for these data). The validation set represents the complete set of 'discipline' degree materials (detailed in table 13, section 2.4.1). For the training set we randomly selected 1,000 job adverts from each discipline, with the remainder used to comprise the test set.

The next step was to optimise the hyperparameters; performed in two stages using grid search. Firstly, we optimised the parameters for the Support Vector Classifier (C and the tolerance). For the penalty parameter (C) we tested the values of 1.0, 0.5 and 0.1, and for tolerance 0.001, 0.0001 and 0.00001. The values that performed best were $C = 1.0$ and tolerance = 0.0001, which are the default values for the package, further suggesting their validity. For the bagging extension, we optimised for the number of estimators (classifiers), and the number of samples to be drawn. For the former we tested for 10, 100, 250 and 500 estimators, and for the number of samples we drew both 100% and 50%. The best performing parameters were for 100 classifiers and with 100% of the data used for sampling.

Having identified the optimal parameters for the model, we then used it to predict the test data. The model was run with an overall accuracy of 0.638. Whilst this figure is less than desirable, identifying the correct class in just under two thirds of cases, this is not necessarily surprising. We may rightly raise concerns about data quality, not only as this is text data extracted from the internet, but also, we may well expect some overlap between classes, both with the job adverts and course materials. Ultimately, however, the level of accuracy required is dependent on the purpose of the classifier and the data used. In this case, we are simply looking for an indication of similarity between materials, and therefore the stakes are far lower than in, for example, healthcare or credit domains. Held in contrast to a completely random, unbiased classifier, which would on average correctly assign the class at a level of 1 / the number of classes, our classification level of 0.638 outperforms the random figure (0.167) by a ratio of nearly 4 to 1.

Further insight to the performance of the classifier can be gained by analysis of the confusion matrix, shown in table 25. Through identifying where classes have been repeatedly misassigned, we can evaluate specific classes that have reduced this average and increased inaccuracy.

Table 25 Confusion matrix for the test data (job adverts)

Actual Class	Predicted Class						
	CS	IS	ML	OR	PS	ST	Total
CS	276	102	55	14	8	21	**476**
IS	97	310	7	18	35	27	**494**
ML	37	6	299	29	25	19	**415**
OR	28	56	27	248	37	85	**481**
PS	3	11	3	19	447	17	**500**
ST	31	34	30	92	56	236	**479**
Total	**472**	**519**	**421**	**420**	**608**	**405**	**2,845**

CS = Computer Science; IS = Information Systems; ML = Machine Learning;
OR = Operational Research; PS = Psychology; ST = Statistics

As table 25 demonstrates, there are some cases where the classifiers have particularly struggled to distinguish two of the classes. In 102 instances CS adverts were misclassified as IS (21.43%), and 97 instances where IS adverts were classed as CS (19.64%). Although less pronounced, there was a similar issue distinguishing statistics and OR. There were 85 instances where OR adverts were classified as ST (17.67%) and 92 where ST adverts were classified as OR (19.21%). Whilst this is clearly detrimental to our confidence in the model, this is perhaps not completely surprising. Fundamentally many jobs in the computer science domain would suit IS graduates (and vice versa), and many OR and statistics roles will share considerable communality and overlaps.

To further explore performance, we calculated further metrics. Three of the most commonly used are (where $TP =$ True Positives; $FP =$ False Positives; $FN =$ False Negatives):

$$\text{Precision} = TP \div (TP + FP)$$

$$\text{Recall} = TP \div (TP + FN)$$

$$F_1 = 2TP \div (2TP + FP + FN)$$

In other words, precision gives a measure of the proportion of correct predictions out of all of the predictions made for that class; recall gives the proportion of correct predictions by the total number of instances of the class in the test data; and F_1 gives the harmonic mean of these two metrics. The results of these metrics are shown in table 26. As can be seen, there is some discrepancy between the effectiveness of the classifier for different disciplines. Psychology is the best performing with an F1 score of around 0.8 (mostly due to a recall score of nearly 0.9), whereas computer science, OR and statistics are at the lower end with F1 scores less than 0.6.

Table 26 Precision, recall and F_1 measures for the test data (job adverts)

	Precision	Recall	F1
CS	0.5847	0.5798	0.5823
IS	0.5973	0.6275	0.6120
ML	0.7102	0.7205	0.7153
OR	0.5905	0.5156	0.5505
PS	0.7352	0.8940	0.8069
ST	0.5827	0.4927	0.5339

Overall the model has demonstrated reasonable accuracy and predictive power, and therefore we proceeded to use it to predict the de facto validation set; the course materials data excluding analytics degrees. As stated in the methodology, for the model to be valid we would require no significant drop in accuracy when using course data, particularly considering that some doubts remain about the overall performance using only job advert data. In fact, this data marginally outperformed the job advert data, with an overall accuracy of 0.655. The implication is that the model has grasped much of the uniqueness of the disciplines, and that the consistency of the job advert test data may be the main reason why its predictive power may be lower than ideal.

Accuracy, however, has limitations in when class sizes are imbalanced. Whilst the test data had similar sample sizes for each category, the validation set was determined by what was available online, and therefore there are considerably more examples of computer science, for instance, than OR. For further analysis, and to provide some control for this factor, again the confusion matrix and performance metrics were calculated and are displayed as table 27 and table 28 respectively.

Table 27 Confusion matrix for the validation data (course materials)

		Predicted Class						
		CS	IS	ML	OR	PS	ST	Total
Actual Class	CS	20	4	22	11	12	0	**69**
	IS	0	30	2	2	6	0	**40**
	ML	1	0	11	3	1	0	**16**
	OR	0	0	0	10	0	0	**10**
	PS	0	0	0	0	37	0	**37**
	ST	0	0	0	2	0	17	**19**
	Total	**21**	**34**	**35**	**28**	**56**	**17**	**191**

Table 28 Precision, recall and F_1 measures for the test data (course materials)

	Precision	Recall	F1
CS	0.9524	0.2899	0.4444
IS	0.8824	0.7500	0.8108
ML	0.3143	0.6875	0.4314
OR	0.3571	1.0000	0.5263
PS	0.6607	1.0000	0.7957
ST	1.0000	0.8947	0.9444

Analysis of tables 27 and 28 demonstrates some noticeable improvements over the job advert data. Indeed, OR and Psychology both have 100% scores for recall, and statistics 100% for precision. Statistics' overall records a very high performance now, whilst the biggest concerns are computer science, machine learning and OR. Computer science presents very high precision metrics (over 95%), yet there are several instances where course materials have been incorrectly predicted as alternative classes, particularly for machine learning. Machine learning has reasonable recall (with only 5 courses misaligned), but with so many computer science courses predicted in this class the overall metrics are low. Indeed, a qualitative analysis of some of the materials associated with both does indeed show a great deal of overlap between the two, with many computer science courses featuring machine learning type modules.

There are relatively high numbers of computer science misclassified as psychology and OR, the latter negatively impacting the F_1 scores for OR. This problem is likely to be in part due to smaller numbers of OR examples, in particular. Whilst there is no exact answer to what constitutes an acceptable F_1 score, and by default this is very dependent on the proportions of class size in the data, an option is to compare the score to that of a random classifier. Unlike before, there is no easy calculation for this. However, a figure can be approximated using simulated Monte Carlo trials of a classifier on a dataset with equivalent proportions. In other words, given that 'OR' documents represent 10 out of the total of 191 (approximately 5%), we can simulate a binary random classifier, one which is simply assigning 'OR' or 'not OR', over several trials (arbitrarily set at 10,000), and from this derive F_1 scores. Following this process, the simulation produced an averaged F_1 score of 0.0500, compared to the score of 0.5263 achieved for the OR class. This difference is particularly notable (an improvement of over 1,000%), considering our algorithm has a much harder task as it was classifying across six classes rather than the binary problem given to our random classifier.

We also consider the overall F_1 score of the model. Compared with a score of 0.636 for the test set, the validation set achieves 0.731, which not only represents a notable improvement, but also a score we may consider as relatively healthy given the imbalance in the dataset and the relative lack of precision with which the data has been collected (in comparison to more 'traditional' approaches).

Overall, we conclude the model has been shown to have reasonable predictive power, particularly when considering the F_1 score achieved on the validation set, and that the use of job advert data is a reasonable proxy for course materials (in fact improving the model). However, we do have some remaining concerns, particularly in predicting the computer science class.

The final stage is then to predict the 'analytics' degree materials. Obviously, with no analytics class, there are no accuracy or performance metrics to report, however, we are, of course, interested in the model's class predictions. Of the 43 degrees, 20 were classified as OR; 17 as machine learning; 4 as psychology; 2 as statistics; and none for IS and computer science. The prominence of the OR classification is noteworthy (and is discussed later). There is some concern that computer science is under-represented in this assignment, or more accurately not represented, particularly as the discipline had the lowest recall rate (0.2899). Ultimately there are no easy responses to this situation, and must be considered a limitation of the work. Each of these classifications will be discussed in sequence, with a brief qualitative analysis of associated course materials to validate and further illustrate the characteristics of these classifications.

5.2.1 The Statistics Classification

The two degrees classed as statistics display a very clear connection from the titles alone; Aston University's Business and Marketing Analytics and the University of Edinburgh's Marketing & Business Analysis. Evaluation of the online materials demonstrates that alongside some marketing and business content, both courses also include clear statistical content, for example, with modules around forecasting and market research. Furthermore, it is arguable that the marketing element of these degrees necessitates a greater statistical focus, as many of the analytics performed in marketing fit with traditional statistics (such as surveys and experimental design), and the relationship between marketing and statistics is a long and well-established one.

5.2.2 The Psychology Classification

The four degrees classified as psychology represents the more surprising result of the model, particularly as the module analysis found relatively little connection between psychology modules and analytics modules. However, as psychology had 100% recall in the validation, the result can be presented with some confidence. Seemingly, the degrees in this class are more diverse and the connection to the predicted discipline less apparent. In the first title, Data Analysis, Visualisation and Communication (the University of Aberdeen), a link can be found on the focus on visual and other forms of communication. Similarly, for Swansea University's Management (Business

Analytics) we could argue an emphasis on management may provide a connection to psychology as this represents primarily a human-orientated practice where influencing and communication skills are paramount.

The final two identified in this class, Manchester Metropolitan University's Business Technology and Analytics and the University of Sheffield's Data Science, may seem more unlikely candidates for this classification. However, reviewing some of the materials demonstrates a focus on management issues rather than solely technical, for example:

> "The programme will allow you to develop the acumen necessary to identify opportunities for the organisation through understanding data, as well as the ability to build a case for exploiting and deploying technological solutions." (Manchester Metropolitan University)

> "We can bring you up to speed with the latest technology and management techniques. But we're also deeply concerned with the social, ethical and moral implications of the data revolution." (The University of Sheffield)

In summary, two possible conclusions could be reached. Firstly, this slightly surprising result may simply be due to the limitations of the classifier. However, an argument can also be made that this topic is classified as psychology more because it is focused on the decision-making end of the analytics spectrum than because it features the topics typical of a psychology degree.

5.2.3 The Machine Learning Classification

The second most common classification, machine learning, occurred 17 times (nearly 40%). Whilst there are too many to discuss individually, there are some clear patterns. In terms of degree titles, 7 include the phrase "data science" and 7 the phrase "big data". Indeed, all bar two had at least one of these two phrases (with Nottingham Trent University's Data Analytics for Business and the University of Leeds' Advanced Computer Science (Data Analytics) as the exceptions). Further to this, all of the degrees were based at computing schools (or similar) with the exception of Nottingham Trent's, which is hosted in both the business school and the school of science and technology (the latter of which also houses computer science). This increased role of computer science in this version of analytics is illustrated in the course descriptions of some of the degrees. For instance:

> "Data science is a field of computer science which is concerned with the manipulation, processing and analysis of data to extract knowledge. This area is undergoing a revolution in which HPC [High Performance Computing] is a key driver" (The University of Edinburgh)

5.2.4 The Operational Research Classification

Finally, we considered the materials associated with OR-classified degrees. Again, clear patterns can be seen amongst those classed as such. Firstly, there are several degrees that from the title alone would suggest an OR flavour. Examples include the University of Leeds' Business Analytics and Decision Sciences, the University of Manchester's Business Analytics: Operational Research and Risk Analysis, and the University of Southampton's Business Analytics and Management Sciences. There is also more of a "business" orientation in comparison to the previous class, with 12 featuring that keyword, and other related "business" keywords in their titles, such as Loughborough University's Business Analytics Consulting and Birmingham City University's Data Analytics and Management. However, three of the degrees have "data science" in their title, suggesting that this term is perhaps not purely being used in association with machine learning and computing.

For this classification, there is more variety in the schools which host the programs. The majority are within business schools, 11 in total. However, there are 5 in computing schools, 2 in science and technology, and 2 that are hosted between both the business school and computing school.

5.2.5 Summary

This analysis presented in this section has suggested several insights as to the forms of analytics degrees in the UK. However, in doing so the results suggest there are multiple 'types' of analytics degrees, each of which align to different disciplines. Marketing orientated degrees seemingly differ from more general analytics degrees, particularly incorporating a greater focus on statistical methods. There are also analytics degrees, classed in this analysis as "psychology", which incorporate a greater communication and management orientation. However, the SVC labelled most of the degrees, by a substantial margin, as one of two categories; either based on a similarity to the OR or machine learning disciplines. Consequently, for next stage of the module analysis, which is presented in the next section, we elected to analyse not only the full analytics dataset, but also on the subsets that have been labelled 'OR' and 'ML' in the SVC.

5.3 A Curricula Analysis Using Module Topic Weighting

As detailed in the methodology, the second analysis performed was solely on the module titles contained in the degree materials. Thematic codes were developed and assigned to each module in the traditional discipline set (i.e. all materials other than those associated with "analytics). In total, there were 106 codes identified across nearly 2,000 modules. Using the MTW weighting scheme (presented in the methodology, section 2.5.2) each discipline's codes where given a score that sought to capture the discriminatory value of that topic to the discipline in comparison to the others. Using this scheme, we therefore identified the most important topics by discipline, as presented in figure 22.

Figure 22 Top 10 highest weighted topics (*MTW*) by discipline (in descending order of discriminatory power)

Computer Science	
Term	**MTW**
Graphics	0.5406
Distributed computing	0.5224
Computer architecture	0.5197
Mobile	0.5057
Internet programming	0.4947
Software	0.4649
Computer security	0.4592
Programming	0.4531
Multimedia	0.4304
Networks & servers	0.4249

Information Systems	
Term	**MTW**
Strategy	0.8140
Performance management	0.7821
Enterprise resource planning	0.7804
Management	0.7764
Information systems	0.7415
Knowledge management	0.7094
Project management	0.6897
Business intelligence	0.6632
Operations management	0.6486
Human resources	0.6486

Machine Learning	
Term	**MTW**
Robotics	0.8416
Natural language processing	0.8130
Image processing	0.8045
Machine learning	0.8026
Computer vision	0.7691
Visualisation	0.7581
Artificial intelligence	0.7104
Business intelligence	0.6627
Agents	0.6607
Neuro science	0.5777

Operational Research	
Term	**MTW**
Spreadsheets	0.9420
Supply chain management	0.9171
Operational research	0.9137
Decision sciences	0.9033
Operations management	0.9033
Game theory	0.9033
Optimisation	0.8943
Consulting	0.8937
Simulation	0.8922
Stochastic modelling	0.6713

Psychology*	
Term	**MTW**
Psychology	0.8954
Business psychology	0.8954
Social psychology	0.8954
Cognitive psychology	0.8954
Clinical psychology	0.8954
Neuro science	0.8431
Human resources	0.7211

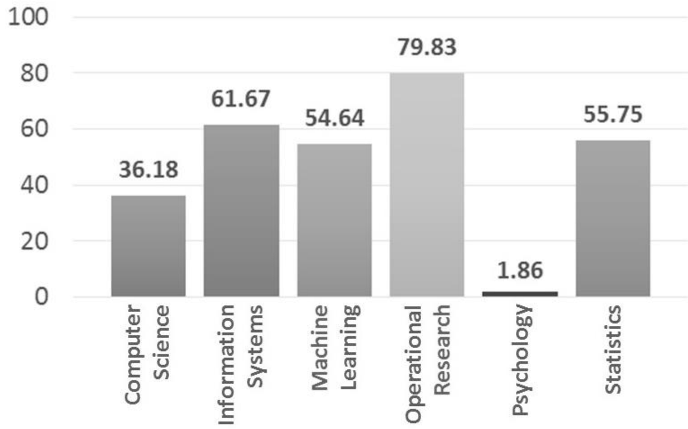
Statistics	
Term	**MTW**
Bayesian statistics	0.8701
Hierachical data	0.8701
Experiments	0.8701
Surveys & sampling	0.8701
Linear models	0.8701
Regression	0.8701
Survival analysis	0.8701
Geospatial	0.8701
Monte Carlo	0.8701
Medical & health	0.8588

**Psychology only includes 7 topics here as the remainder occurred in two degrees or fewer and/or represented less than 1% of the total topics in the discipline, and therefore were ineligible for MTW scores*

Through visual analysis of this list, some degree of 'face validity' is given to the approach; the topics highlighted in each discipline do display relatively strong association with their fields. However, it is important to note that there is a gap between these and the most frequent topics in each discipline (or indeed the topics the casual observer may most associate with each). One obvious example is OR and "optimisation". This topic is indeed the most frequent of all in the discipline, and the one many will most closely associate with OR. However, in figure 22 this is shown to be only 7th in terms of 'importance' using this weighting scheme. The cause is that "optimisation" also features at least once in each other discipline category except for psychology. In other words, if an analytics degree features an "optimisation" module we could not be completely certain that this is indicative of an OR association, merely that there is a strong probability, something which the MTW weighting reflects.

Using these code frequencies, we can compute a total score for each discipline; that is the sum of the frequency of all topics in the full analytics corpus multiplied by the individual weightings of that term in each discipline (the majority of which were zero for the reasons given above). In other words, we provide a probabilistic judgement (based on MTW), as to the extent to which the analytics modules relate to the 'traditional' disciplines included in the study. This allows for a comparison of these totals across disciplines, an indicator of the relevance of each in the curricula of analytics degrees, as shown in figure 23.

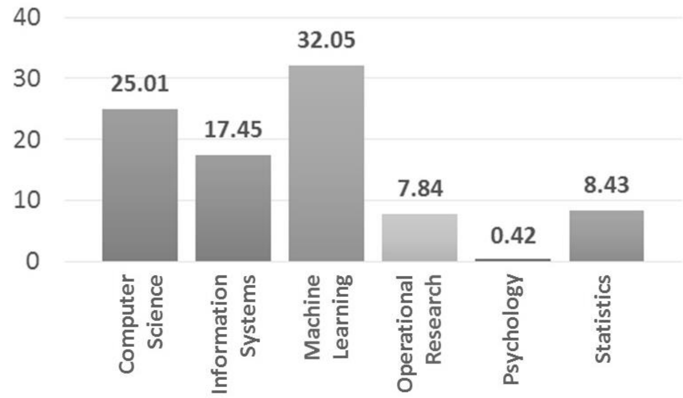
Figure 23 Summed module scores by discipline (analytics degrees)



As with the SVC model of section 5.1, the most prominent discipline in this analysis was OR. In contrast, however, IS is the second most prominent, with machine learning only fourth, scoring marginally below statistics. Psychology scores very lowly, not completely surprising in terms of *a priori* theory, but seemingly contradictory to the result of four degrees being classed with this label. However, whilst these results are useful, the indication of the previous analysis is that, to some degree, this may be comparing apples and oranges, in that the main analytics degrees incorporate two separate categories; those classed as machine learning and another as OR. Therefore, we performed the same procedure on each of these subsets (separating based on classification) to analyse the module weightings associated with each.

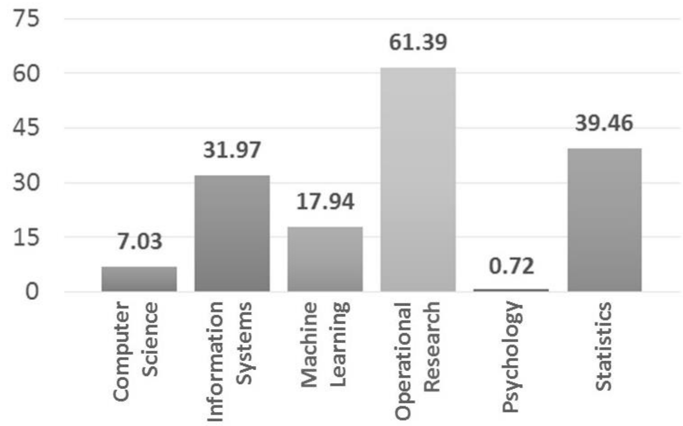
Despite the lower overall score of machine learning, obviously, it would be expected that the discipline would score better in the subset of degrees classified as "ML" in the SVC. As shown in figure 24, the results on this subset alone, this indeed was the case. Machine learning is now by far the most prominent, with computer science the second highest scored. OR is far less influential in this dataset as the second smallest, with statistics taking a significantly lower position than with the totals. IS remains reasonably prominent as the third highest in terms of module topic scores.

Figure 24 Summed module scores by discipline ('ML' classed degrees)



Finally, we analysed the subset that were classified as "OR", shown in figure 25. As would be expected, OR is the most prominent discipline in this dataset, followed by statistics and IS (again taking the third position). Machine learning and computer science drop the most in comparison to figure 23, now ranking 4th and 5th respectively.

Figure 25 Summed module scores by discipline ('OR' classed degrees)



To further analyse each, the top 20 topics in each subset are reviewed, as shown in figure 26. The first category displays, unsurprisingly, a close association to machine learning, and related fields such as big data, cloud computing and data mining and specific techniques such as visualisation and natural language processing. Additionally, there is a clear emphasis on computer science-type topics including programming, security, databases and networks & servers. The second category features many of the topics prominent in OR courses such as forecasting, optimisation and simulation, whilst statistics is listed as the third most frequent topic. There is also a very clear business theme, with a variety of domains including finance and marketing (the top two) as well as more general business topics such as management, strategy and operations management. As discussed, decision making modules are more prevalent here, with decision sciences and consulting key topics (additional to the domain specific terms already discussed). The module topics which rank highly in both include "data mining", "data management" and "statistics".

Figure 26 Top 20 topics in the two categories of analytics degrees (by frequency)

First Category (ML Classed)		
Rank	**Module Topic**	**Freq.**
1	Big data	15
2	Data mining	9.5
3	Machine learning	8.5
4	Web & eBusiness	8
5	Programming	7.5
6	Data management	5.5
=	Natural language processing	5.5
=	Visualisation	5.5
9	Cloud computing	5
=	Computer security	5
=	High performance computing	5
=	Statistics	5
13	Computer science	4.5
=	Data analysis	4.5
=	Databases	4.5
=	Information retrieval	4.5
17	Distributed computing	4
=	Networks & servers	4
19	Optimisation	3.5
=	Software	3.5

Second Category (OR Classed)		
Rank	**Module Topic**	**Freq.**
1	Finance	14.5
2	Marketing	11
3	Statistics	10.5
4	Data analysis	9.5
=	Data mining	9.5
6	Decision sciences	8
7	Supply chain management	7.5
8	Management	7
=	Strategy	7
10	Forecasting	6.5
11	Consulting	6
=	Data management	6
13	Web & eBusiness	5
=	Operations management	5
=	Optimisation	5
=	Operational research	5
=	Project management	5
=	Simulation	5
19	Economics & econometrics	4.5
20	Machine learning	4
=	Natural language processing	4

5.3.1 Summary

The module analysis presented in this section was designed to evaluate the specific skills, subjects and techniques taught within analytics degrees. However, following on from the findings of the previous section (classification of degrees), this analysis is seemingly more meaningful when applied to the two main categories of analytics degree found in the earlier analysis; the degrees which most closely align to machine learning, and to OR respectively. The two categories demonstrate a relatively 'clean' separation. The skills and techniques most frequently taught in each not only align to these disciplines (machine learning and OR), but also demonstrate other unique characteristics such as an association with computing in the case of the former, and business topics with the latter. Such a separation provides further validation to the SVC analysis, but also provides further detail on the different characteristics of each, and a more granular perspective on the specific approaches common to each. This line of thought is concluded in the following section, where a more complete description of the two 'types' of analytics degree these analyses suggest.

5.4 A Typology of Analytics Education

The analyses, presented in sections 5.1 and 5.2, suggests that, in combination with other hybrid-type degrees that seemingly combine analytics with management or marketing topics (and indeed other specialisations would be feasible), analytics master's degrees in the UK broadly fit into two categories. The first category is most closely aligned to machine learning, primarily emerges from computing and technology schools, whereas the second category, aligned to OR, will typically be based in business schools.

These findings are supported by analysis of universities that offer more than one degree in the analytics area. In total 7 universities fall into this category as shown in table 29. In most of these cases there is a clear 'two-pronged' approach, with a business school-based course classified as OR, often named a variant on "business analytics"; and a course with a classification of machine learning, based in computing schools, and typically with "data" in its name. There is one surprising results, Lancaster University's courses in 'Data Science' and in 'Management Science and Marketing Analytics' both carrying a classification of 'OR'. However, aside from this the pattern is relatively clear.

Table 29 Universities with multiple analytics-type degrees

University	Degree Title	School	Class
University of Edinburgh	Marketing & Business Analytics	Business School	ST
University of Edinburgh	High Performance Computing with Data Science	Parallel Computing Centre	ML
University of Essex	Business Analytics	Business School	OR
University of Essex	Big Data and Text Analytics	Computer Science & Electrical Engineering	ML
University of Essex	Data Science	Computer Science & Electrical Engineering	OR
Lancaster University	Data Science	Science & Technology	OR
Lancaster University	Management Science and Marketing Analytics	Management School	OR
University of Leeds	Business Analytics and Decision Sciences	Business School	OR
University of Leeds	Advanced Computer Science (Data Analytics)	Computing	ML
Swansea University	Finance and Business Analytics	School of Management	OR
Swansea University	Management (Business Analytics)	School of Management	PS
UCL	Business Analytics	Computer Science & Management Science	OR
UCL	Web Science & Big Data Analytics	Computer Science	ML
University of Warwick	Business Analytics	Business School	OR
University of Warwick	Data Analytics	Computer Science	ML

There are several possible reasons as to why this separation may occur. In an ideal world, the assumption would be that the different degree categories have emerged to meet specific training needs, such as those discussed at the start of the chapter. However, it also appropriate to acknowledge that there may be more pragmatic reasons behind this. The provision of master's degrees is obviously dependent on teaching resources, and it is obviously easy, quicker and less costly for a university to utilise existing staff to this end. In the case of the creation of analytics degrees, if there are staff already employed with experience in areas such as machine learning and OR, this could explain why degrees come to take these attributes. Secondly, most universities will be divided into specific schools and faculties, each of which have their own specialisations and topic boundaries. If, as this research would suggest, analytics has aspects of both computational elements as well as business elements, the school in which the degree is hosted is likely to have strong bearing on which of these orientations is stronger in its curricula.

Regardless of cause, there seems relatively strong evidence in the data to support the categorisation, and enough 'distinctiveness' in each to draw some characterisations. The characteristics of each of these are presented as a summary in table 30, whilst the remainder of this report will consider these implications, and also the contributions of this research.

Table 30 Summary characteristics of the two types of analytics degrees

	Type One	**Type Two**
Likely title(s)	"Data Science" / "Big Data Analytics"	"Business Analytics"
SVC classification	Machine learning	Operational research
Likely school	Computing / Technology school	Business school
Linked disciplines	Machine learning (35.15%) Computer science (27.42%) Information systems (19.13%)	Operational research (38.72%) Statistics (24.90%) Information systems (20.17%)
Most likely module topics	Big data (9.46%) Data mining (5.99%) Machine learning (5.36%) Web & eBusiness (5.05%) Programming (4.73%)	Finance (7.20%) Marketing (5.46%) Statistics (5.21%) Data analysis (4.71%) Data mining (4.71%)

5.5 Discussion and Implications

The results of the analyses presented in this chapter, suggesting two seemingly quite contrasting types of analytics degrees are currently offered at UK universities, has significant implications for the research, as well as for analytics, OR and the variety of related disciplines. Firstly, again, the taxonomy of figure 14, section 2.3, and therefore our conceptualisation of the disciplines inherent to analytics, has been further validated. This is particularly demonstrated in the classification of four analytics degrees as most closely aligned to the psychology discipline (as a part of the "decision making" element of analytics). This aspect was not as clearly seen in the analyses of chapter four.

Secondly, and in keeping with the results of chapter four, we see a clear alignment between analytics and machine learning. Indeed, the results of this chapter would suggest that the machine learning discipline may have a far greater role in analytics than previously suggested; indeed, to the extent that it represents direct competition to OR for a 'share' of the field. In comparison to chapter four, statistics is far less represented in the SVC model, however, shows a key relevance in the module analysis for the degrees classed as 'OR'.

5.6 Summary

In summary, this chapter has sought to classify analytics degrees to the ‘traditional’ disciplines they are most similar to, firstly on a general level, and then at a module level. The results have suggested a ‘two-tiered’ approach with degrees that show marked similarity to OR degrees, and typically based in business schools, as well as degrees relating to machine learning, most frequently in computing or technology schools. Whilst the research purposively sought to avoid distinguishing between analytics and data science, as there are many sizeable overlaps between each and ambiguities as to where one starts and ends, the results of this analysis problematise such as stance. In essence, the implications are that, in university master’s degrees at least, there is a distinction made between the two, with data science having stronger ties with computer science, machine learning and big data; whereas analytics links more closely to business, to domains such as finance, and to OR. This line of thought will be continued in the discussion (chapter eight).

These results provide important insights towards addressing RO4, identifying the disciplines with which analytics degrees align most closely with, which will be verified and explored further in chapter seven, the interviews with academics and course developers. However, prior to this, the next chapter will report the results of the interviews with analytics and OR practitioners and employers.

6 Hire Education: Employer Skills Requirements for Analytics Graduates

In chapter four an initial (quantitative) analysis was performed towards meeting the third objective of the research, effectively determining the requirements made of graduates for work in the analytics and related roles. As is typical of the instruments, the quantitative analysis (chapter four) provided a relatively *broad* analysis, summarising the more 'macro' differences between analytics jobs and those aligned to more traditional disciplines, from which we can make some inferences about what is 'new' in the requirements of analytics roles. To complement this analysis, and to add a greater *depth* to our conceptualisation of the requirements of analytics roles, this chapter presents the results of a qualitative analysis of 29 interviews with OR and analytics employers (and potential employers).

To this end, the chapter is organised as follows. We begin by presenting the template used for the analysis with a brief discussion on how it was developed. Thereafter, the main topics of the template are discussed in sequence; incorporating data management, quantitative methods, IT and soft skills, internal processes, analytics outputs, education and future trends. This is followed by a matrix analysis, from which specific examples will be analysed in greater depth, before, finally, a brief analysis and discussion of the results is provided.

6.1 Research Sample

As detailed in the introduction and earlier methodology, in this part of the research a sample size of 29 were recruited from a range of company types (a full listing of participant counts per category is given in chapter two, section 2.6.1). Participant names are omitted to protect their privacy, but each is referenced with a unique name related to their occupation. A full list of names, company types and categories is shown in table 31.

Table 31 List of interview participants with coded name, company type and category

Interviewee	Company Type	Category
Technology Consultant	*Consultant*	*Consultant*
Analytics Manager (Telecoms)	*Utilities*	*Employer*
Digital Analytics Consultant	*Marketing*	*Consultant*
Government Analytics Manager	*Public*	*Employer*
Analytics Consultant (Smaller Management Consultancy)	*Consultant*	*Consultant*
Analytics Manager (Health)	*Public*	*Consultant*
Analytics Manager (Utilities)	*Utilities*	*Employer*
Analytics Consultant (Finance)	*Consultant*	*Consultant*
Analytics Manager (Online Travel)	*Travel*	*Employer*
OR & Analytics Consultant	*Consultant*	*Consultant*
Government Data Scientist	*Public*	*Employer*
Analytics Manager (Public)	*Public*	*Employer*
Marketing Analytics Consultant	*Marketing*	*Consultant*
Software Vendor (Data Management)	*Vendor*	*Vendor*
Government Analytics Manager (Finance)	*Public*	*Employer*
Software Vendor (Analytics General)	*Vendor*	*Vendor*
Analytics Manager (Consultancy)	*Vendor*	*Vendor*
Analytics Manager (Energy)	*Utilities*	*Employer*
Analytics Consultant (Larger Management Consultancy)	*Consultant*	*Consultant*
Analytics Manager (Management Consultancy)	*Consultant*	*Consultant*
Software Consultant (Simulation - Processes)	*Vendor*	*Vendor*
Marketing Analytics Manager	*Marketing*	*Consultant*
Analytics Recruitment Consultant (Niche)	*Recruitment*	*Recruitment*
Analytics Manager (Retail Travel)	*Travel*	*Employer*
Analytics Recruitment Consultant (Larger)	*Recruitment*	*Recruitment*
Software Vendor (Simulation - All)	*Vendor*	*Vendor*
Healthcare Analytics Consultant	*Vendor*	*Vendor*
Software Vendor (Information Technology)	*Vendor*	*Vendor*
Media Company Analyst	*Media*	*Employer*

Each interview was recorded, transcribed and then analysed using template analytics and matrix analysis. The remainder of this chapter will discuss the results of these analysis.

6.2 Template Analysis

The first step in analysing the interview transcripts, involves building an initial template. Initially, a broad set of *a priori* codes were created, as shown in figure 27, which roughly correspond to the primary question set (described in section 2.6.2). These codes are conceived as the ‘parent nodes’ in a hierarchical coding structure, from which the ‘child nodes’ can be inferred from the data itself (as can new ‘parent nodes’). Equally, however, nodes too can be removed, combined or mutate depending on the level of support in the data.

1. Data
2. Quantitative methods
3. Programming & software
4. Soft Skills
5. Other skills
6. Education
7. Outputs
8. Internal organisation
9. Recruitment
10. Analytics trends

Figure 27 Initial template ‘parent nodes’ in order of interview topics

To develop the initial template, a subset of four transcripts, drawn from each of the main categories of interviewee type (two ‘employers’; two ‘consultants’; one ‘software vendor’; and one ‘recruitment consultant’) was analysed with the approach listed above. In doing so, a full initial template was built, and is shown in figure 28. Most of the ‘child nodes’ added were expected, for instance “big data”, “OR” and “data visualisation”. However, some were less so, for instance the importance of “maths pre-degree”, and an emphasis on “design of experiments” in one interview. Also, as can be observed, the template’s hierarchy goes to a third-level of depth.

Having established the initial template, the remaining interviews are analysed against it. As the analysis progresses, not only were the transcripts coded, but also the template refined. In most cases, the goal was towards reductionism and the simplification of the template. This was predicated on one of two rationales. Firstly, on a basis of ‘outlier control’. Some topics were relevant to the four test cases on which the initial template was based, but were common in subsequent interviews. An example of this was “design of experiments”. Whilst this something that is clearly relevant to analytics, its lack of prominence across the whole dataset led to the decision to subsume it within the wider, second-level topic of “statistics”. Secondly, in some cases a code was found to “substantially overlap with other codes” (King, 2004, p 262) such that it made more sense to combine them. Ultimately, this was the decision taken for the previously distinct codes of “SAS”, “R” and “SPSS” (into a code of “statistical languages and software”). Equally, much of the talk about recruitment, also related to specific skills. Rather than keep this as a distinct ‘parent node’, and to avoid duplication, these areas are discussed alongside the underlying skills requirements. The final template is presented in figure 29.

1. Data
   - 1.1 Ad-hoc data
   - 1.2 Big Data
     - 1.2.1 High volume data
       - 1.2.1.1 Not the issue
     - 1.2.2 High velocity data
     - 1.2.3 Unstructured data
   - 1.3 Data management
     - 1.3.1 Data quality & cleaning
     - 1.3.2 Data strategy
   - 1.4 Traditional data
2. Quantitative methods
   - 2.1 OR
   - 2.2 Statistics
     - 2.2.1 Design of experiments
   - 2.3 Machine learning
   - 2.4 Data mining
     - 2.4.1 Importance of *a priori*
   - 2.5 Economics
3. Programming & software
   - 3.1 Bespoke solutions
   - 3.2 Big data languages
   - 3.3 High-level languages
   - 3.4 SAS
   - 3.5 R
   - 3.6 SPSS
   - 3.7 VBA
   - 3.8 Web languages
   - 3.9 SQL
   - 3.10 OR tools
   - 3.11 Understanding principles
   - 3.12 Programming unimportant
4. Soft Skills
   - 4.1 Communication
   - 4.2 Are crucial
   - 4.3 Relationship building
   - 4.4 Feedback & reflection
   - 4.5 Handling objectives
   - 4.6 Influencing
   - 4.7 Presenting
   - 4.8 Problem structuring
   - 4.9 Soft systems methodology
   - 4.10 Sales skills
   - 4.11 Translation
   - 4.12 Understanding customer/brief
   - 4.13 Visual design
5. Other skills
   - 5.1 Change management
   - 5.2 Domain experience
   - 5.3 Six-sigma and lean
   - 5.4 Strategy & value management
   - 5.5 Project management
     - 5.5.1 Agile
     - 5.5.2 Prince 2
     - 5.5.3 Waterfall
     - 5.5.4 As an aptitude
     - 5.5.5 General appreciation
     - 5.5.6 Separate PMs
   - 5.6 Structuring analytics in the business
   - 5.7 Analytics mindset
6. Education
   - 6.1 Degree
     - 6.1.1 Any quantitative
     - 6.1.2 Computer science
     - 6.1.3 Maths & statistics
     - 6.1.4 OR
     - 6.1.5 Sciences
     - 6.1.6 Specialised analytics
     - 6.1.7 Other disciplines
   - 6.2 Difficulty testing soft
   - 6.3 Maths pre-degree
   - 6.4 Online qualifications
     - 6.4.1 Not credible
     - 6.4.2 Not useful in quants
     - 6.4.3 Technology
   - 6.5 Presentations
   - 6.6 Projects
   - 6.7 Real world experience
     - 6.7.1 Consultancy projects
     - 6.7.2 Internship
     - 6.7.3 Real problems & data
   - 6.8 Red tape causes blocks
7. Outputs
   - 7.1 Data outputs
   - 7.2 Data visualisation
     - 7.2.1 Communication tool
     - 7.2.2 Discovery tool
     - 7.2.3 DIY
     - 7.2.4 Graphs & charts
     - 7.2.5 Not yet established
     - 7.2.6 QlikView/ Tableau
   - 7.3 Dashboards
   - 7.4 Operationalisation
   - 7.5 Presentations
   - 7.6 Reports
   - 7.7 Working models
   - 7.8 Workshops & training
8. Internal organisation
   - 8.1 Data scientist unrealistic
   - 8.2 End-to-end process
   - 8.3 Hard & soft rare
   - 8.4 Multi-skilled teams
   - 8.5 Separate teams
   - 8.6 Teams for big; individuals for small companies
   - 8.7 Technology outsourced
   - 8.8 Analytics domains
     - 8.8.1 Credit & Financial
     - 8.8.3 Marketing/customer
     - 8.8.4 Operations
     - 8.8.5 Sales
     - 8.8.6 Technology
     - 8.8.7 Web
     - 8.8.8 Other
     - 8.8.9 Selling to internal
9. Recruitment
   - 9.1 Experienced hires
   - 9.2 Graduates
     - 9.2.1 Graduate scheme
     - 9.2.2 Greater risk
     - 9.2.3 Masters
   - 9.3 Internal recruitment
10. Analytics trends
    - 10.1 Growth
    - 10.2 Hype
    - 10.3 Increased status of data
    - 10.4 Real-time a fad
    - 10.5 Senior management buy-in

Figure 28 Initial template

1. Data and data management
    - 1.1 Ad-hoc data
    - 1.2 Big Data
        - 1.2.1 High volume data
        - 1.2.2 High velocity data
        - 1.2.3 Unstructured data
    - 1.3 Data management
        - 1.3.1 Data quality & cleaning
        - 1.3.2 Data strategy
    - 1.4 Traditional data
2. Quantitative methods
    - 2.1 OR
    - 2.2 Data mining
    - 2.3 Machine learning
    - 2.4 Statistics
3. Programming & software
    - 3.1 General purpose languages
    - 3.2 Big Data languages
    - 3.3 Statistical languages & software
    - 3.4 Databases & query languages
4. Soft Skills
    - 4.1 Project scoping
    - 4.2 Understanding the customer
    - 4.3 Communicating results
        - 4.3.1 Communication skills
        - 4.3.2 Influencing
    - 4.4 Consulting & sales skills
5. Other skills
    - 5.1 Project & change management
    - 5.2 Domain experience
    - 5.3 Strategy
    - 5.4 The 'analytics mindset'
6. Education
    - 6.1 Disciplines
    - 6.2 Soft skills
    - 6.3 Real world experience
    - 6.4 Other training
7. Outputs
    - 7.1 Data visualisation
    - 7.2 Models & data outputs
    - 7.3 Written outputs
    - 7.4 Workshops & training
8. Analytics trends
    - 8.1 Growth of analytics
    - 8.2 Growing status of data
    - 8.3 Management buy-in
9. Internal organisation
    - 9.1 The 'devolved' approach
    - 9.2 The 'unicorn' approach
    - 9.3 The 'problem solving' approach
    - 9.4 The 'operationalisation' approach
    - 9.5 The 'technical + business' approach

Figure 29 Final template

Such a reductionist approach is not necessarily typical of template analysis, but serves the purposes of this research reasonably well. Whilst, as detailed in section 2.1, it is recognised that interview results cannot be truly generalisable, and obviously, will lack statistical significance of any kind, there is a desire to use these results, alongside the quantitative analysis of chapter four, to identify common patterns and trends that can used to make general recommendations. In part, we would argue for *transferability* as opposed to the *generalisability* that may be found in purely quantitative work (e.g. Marshall, 1996). Beyond this, however, the reductionism described, ensures the result reported have a greater breadth, even at the cost of some of the depth.

6.3 Results of the Template Analysis

Having arrived at the template, and the coding structure it describes, the transcripts were given a final pass, and coded accordingly. The final concern is the presentation of the results. King (2004, p 268) describes three approaches to this, including "an account structured around the main themes identified", which is the approach used for this. As he points out, the main drawback can be "drifting towards generalisations, and losing sight of the individual experiences from which the themes are drawn" (ibid., p 268). This danger, however, will be somewhat countered by the application of a matrix analysis which follows these discussions. Before this, however, each of the topics of the final template, as depicted in figure 29, are discussed in sequence beginning with *data and data management*.

6.3.1 Data and Data Management

As had been indicated when discussing the key events that have led to the development of the analytics period (chapter three), data and data management has been a prominent theme. Naturally, therefore, the topic of data and data management was an important aspect of the interviews and their subsequent analysis. However, within this code we identified several other prominent sub-codes, as illustrated in figure 30.

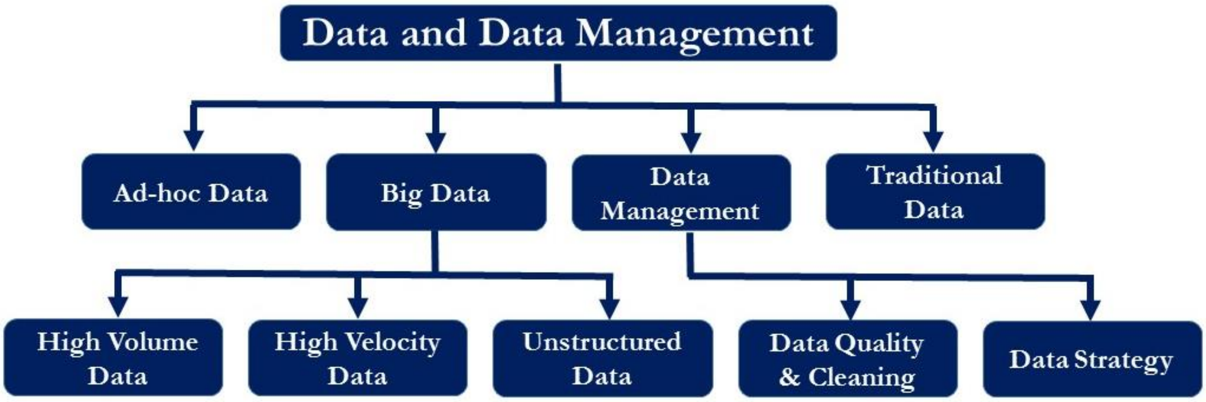


Figure 30 'Data and data management' code hierarchy

Much discussion of analytics concerns *big data* and the vast quantities of information available to organisations in the internet-connected world. However, in discussion with the interviewees, it became clear that there are several situations where *big data* was either unrequired or inaccessible. Often this included scenarios where the main requirement was for *ad-hoc data* collected for a new project, and one where data was not held in company databases nor publicly available. This point is exemplified by an answer given by the manager of an analytics team working on government projects:

> **Analytics Manager (Public):** "Obviously, we've got all our internal data warehouses and we're trying to reconfigure that at the minute, make better use of management information [but] it's much more ad hoc, to be honest [...] We deliver an analytical solution for a project and then we move onto something else, so it's not that kind of environment."

Similar answers we given by consultants and software vendors working on projects. Ultimately in these cases it is the client who provides the data, if indeed it does exist, and therefore the data will be of the form that the client provides, and more than likely this will be traditional data sources rather than big data sources. Further, often for the consultant the original source of the data to be used (be it *big data* or otherwise) is often irrelevant as by the time it is to be used in consultancy projects it has already been cleaned and transformed.

> **OR & Analytics Consultant:** "So, I'm not saying we've got nothing to do with [big data], we're kind of just listening in [...] But there still is going to be data in a database [for] our models at the end of the day."

Even aside such project-based analytics applications, in general most interviewees reported to primarily use "small" data sources. A recruitment agent specialising in analytics reported that whilst they are "seeing more [jobs in big data] our biggest demand is still in the structured data side". However, many of the interviewees were keen to explore this area further, with several about to engage in testing of new data storage structures (such as Hadoop and Spark) or hiring data scientist to join their teams.

The use of such traditional data does of course still make demands on certain skills for employees. Most interviewees, particularly those who were considered direct employers rather than software vendors or consultants, had requirements for databases skills or at least, in the words of a Government analyst, the ability "to understand databases and to be able to develop a data structure". For many this was becoming increasingly important, and that increasing their team's competencies in this area could potentially have a transformative effect on the impact of their work.

> **Analytics Manager (Telecoms):** "We recruited the computer science people to our team 2-3 years ago and they really changed the way we work. We didn't have our own databases before, and without it we'd be a lot more dependent on a sort of business intelligence teams to give us that data. [...] I don't think we quite recognised what [...] a revolution that would bring. [...] Four years ago, I wouldn't have said we need database skills."

Indeed, skills associated with *data management* were a key theme in several interviews. Whilst in most organisations examined there was a specific database team, there was a disparity between the extent to which this resource was relied upon solely, and those who organised additional databases within their team to complement this, such as in the example above. For the former perspective, one interviewee suggested database work is not necessarily suited to analytically trained staff:

> **Analytics Manager (Energy):** "We try to hold [relevant data] in a central repository so we know where that data is, but clearly when you want to look at each particular segment or whatever, or particular project, you'll need to require the data for that. But most of the analytical people, I think, sort of sit in the world where they just specify what they want from that Database Team, rather than [gather it themselves]."

In contrast, for another organisation examined, this issue was so important it necessitated a departmental restructuring:

> **Analytics Manager (Management Consultancy):** "We've taken two teams that used to be separate, so one who specialised in analytics and predictive modelling and another who were technology data structure specialists, and we've combined them together in order to make sure that […] we've got in one place all of the skills we need."

Two issues in *data management* were identified by interviewees as being particularly relevant. Firstly, many companies had sought to create a *data strategy* to maximise the opportunities presented in the analytics period. In many ways, this represents a significant departure from how businesses may have managed data in the past. For several of the organisations examined, data is no longer a resource that is collected purely in an ad-hoc fashion, nor something that is simply stored as an output from business processes in the hope that it may prove useful at a later stage. Indeed, in one company the *data strategy* is of such central importance that it influences which employees are recruited and how analytical teams are structured:

> **Media Company Analyst:** "[The department manager did] a lot of work with the business trying to identify what data is around, what data might be beneficial. And by mapping out where he saw the team going he then worked out what skills we would need to have."

Secondly concerns were expressed about *data quality and cleaning*. Many reported they were increasingly working with messier data in both *traditional* and *big data* projects:

> **Government Data Scientist:** "The data we get is far messier than pretty much anything, well certainly at university all the data I got was lovely. Whereas the data I have now is not lovely at all. Even if the data set is full you have to know where the data is coming from, how much trust you can put in it, how often it's updated, what method it's updated with, all this stuff that certainly when I was at university we didn't think about."

> **Utilities Analytics Manager:** "[The] multiplicity of data that's coming in, and the fact that we're now able to link data from so many different sources in different ways in our data, means that our data governance is much more crucial than it was before, and people in the business need to understand it more than [it] just being an IT thing."

Big data concerns though are not just about *data quality*. Whilst, as discussed, *traditional data* remains the most used by companies in our sample, *big data* is certainly a growing concern in almost all interviews. Following Laney (2001), we use the popular '3 V's' representation of this area (despite the reservations expressed about it in chapter one). For the first of these, *volume*, the consensus was that this is becoming less of an issue. Indeed, many felt that volumes of data naturally do increase, and whilst there has been a technological shift to meet the scale of data currently available, in many ways this remains business as usual.

> **Software Vendor (Data Management):** "For us it's not the volume side, because we can point to ... we had the first terabyte data warehouse, the first petabyte warehouse was from [company], we have people like [online retailer] who have tens of petabyte of data in [the cloud] and I think they process something like 50 terabytes of data a day, they load into their warehouse, which is way more than most of our customers have in their complete warehouse. So, volume isn't the issue."

However, one area within which *data volume* does present new challenges does relate to these new technologies, and specifically finding staff trained in the new software and frameworks associated with *big data*. Indeed, just being able to identify if a potential recruit has the necessary skills can present a challenge in itself.

> **Analytics Recruitment Consultant (Larger):** "It's all very well having the hype for big data skills but if you're a company out there and you want to recruit all these Hadoop people, how are you going to know if they're real or not? What questions can you possibly ask them to actually test out whether they've got the skills which will be helpful to you in what you want and which show that they have got capability?"

This, however, may only be a short-term issue as training and education programs begin to incorporate coverage of these tools into their curricula, either as additions or replacements:

> **Software Vendor (Data Management):** "There is so much hype and visibility around big data that all kinds of people are learning skills and learning technologies [...] the foundation skills around the technologies will become pretty common. Not that they'll not be needed but they actually will just become core skills."

Equally, many of the interviewees called into question concerns about the *velocity* of data. In some organisations applying analytics in real time was not seen as a necessity, particularly in respect to project work which is typically either performed *post haste* or to support strategic decision making (and therefore unlikely to be in 'real-time'). In the words of one of the software vendors interviewed, "real time doesn't necessarily apply in analytics, it's about right time". However, one trend which increasingly means using fast moving data, is in the operationalisation of models.

> **Analytics Consultant (Smaller Management Consultancy):** "A trend I see is the embedding of analytics into an ongoing business process. The days when you used to do an interesting study and produce a report, and not all the time the analytics is embedded into the business process [...] A retailer doesn't produce a report every morning of its sales, but it's got it programmed in so it drives all its restocking from all its stock control. We just think that's normal. In other areas, I think we will see more and more of that."

The final sub-category, *unstructured data*, by contrast is seen to be of greater importance. These data differ from *high volume* and *velocity data* in that they necessitate new quantitative approaches, which has implications both on the methods selected, and the demand for skilled staff.

> **Software Vendor (Data Management):** "[Its] new kinds of data, new sources of data, which are ... and we don't like the word 'unstructured' [...] What they are is they're not traditional relational structures, so they're non-relational structures of data – often things like web logs or machine data or text files – that absolutely have structure. But it's harder to unravel the structure from them. You can't use traditional techniques."

Aside from the technical skills these data necessitate, there are also requirements for staff to be able to make judgements on the veracity of the output and the implications of erroneous conclusions. In contrast to recommendations drawn from traditional approaches performed on highly structured and verified sources (such as financial data), these new sources of data may contain spurious correlations or insights which have limited value to the organisation.

> **Utilities Analytics Manager:** "There are so many of these but nobody knows really whether they're there or how many there really are there. So, then it's about looking at what other sources of data have you got, how can you correlate those sources of data, what kind of match can you get, what kind of mismatch can you get, and therefore what kind of tolerances you have, and what scenarios you might play out."

For these reasons, some organisations who were more frequently using *big data* sources regard them as more useful for experimentation, exploration and for identifying new areas of interest:

> **Software Vendor (Data Management):** "[What] is different in big data, compared to traditional analytics, is the concept that most things won't work. So, you'll have ideas and you'll have hypotheses that you want to test, but [online retailer] quote themselves as saying 80% of the analytical ideas they have, have no value but it's trying to get to the 20% that have value. So, the concept they have of 'fast fail', which is to say you want to try and quickly get through the 80%. And that's what the whole discovery thing is about, saying "bring in new sources, add them to your existing sources, try new techniques, see if they add any value, and then keep chopping and changing."

In summary, and as may be expected, issues surrounding data and its management are of increasing importance to most interviewees. Whilst certainly this does not necessitate that every analysis is one based on big data, there are clear demands for recruits to be data savvy, and with the ability to be increasingly proficient with databases and associated tools.

6.3.2 Quantitative Methods

Drawing again from the taxonomy presented in figure 14 (section 2.3), another major concern were the skills required to perform the quantitative analyses required in the organisation's analytics processes. As before, discussion in this are led to the creation of several sub-categories as shown in figure 31.

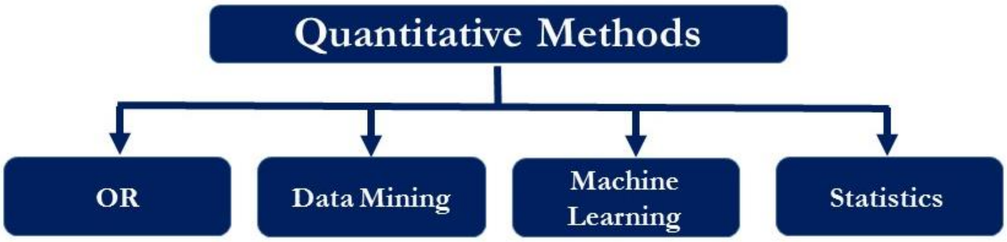


Figure 31 'Quantitative methods' code hierarchy

Our first area of enquiry, and one which is fundamental to our central question, is the use of *OR* techniques. Reassuringly, though perhaps not unsurprisingly considering that most our sample had some affiliation to *OR* (most frequently in their educational background), *OR* skills and techniques were seen to be relevant in both the interviewee's operations and in analytics as a whole. Many had recently employed graduates from *OR* degrees, and/or were involved with providing internships or projects for students to engage with during their study. Indeed, for one of the participants there had been a significant growth in demand for *OR* graduates and its influence in their (Government) department:

> **Government Analytics Manager:** "We've seen a growth in the number of analysts and the number of analytical disciplines in [the department] since it was formed [...] There were two or three operational researchers, there are now 21 [...] Operational researchers are pretty good at doing these kind of things, so [we're] quite happy having people like that."

In respect to the use of specific techniques, the most frequently listed were optimisation, simulation and forecasting. Agent-based modelling was seen as a key emerging tool for two of the interviewees, but for another two it remained more an interesting concept than something in regular usage, deemed by a Healthcare Consultant to be "very much in academia". Beyond the specific techniques, however, interviewees highlighted the importance of the discipline's focus on problem structuring, model building and the 'way of thinking' associated with the discipline.

> **Software Consultant (Simulation - Processes):** "[O]ur most successful people going back would probably be people who've done a first degree in [a quantitative subject], and then done a Masters in OR. Saying that, given that we focus on simulation, a lot of the other OR techniques [they have] when they come to us, they tend not to use them. But it's a way of thinking."

For interviewees who had less clear links to *OR* though, its utility in modern analytics was more uncertain. Two of the interviewees were not aware of the discipline by name, albeit this may not necessarily mean its methods were not used or that their staff were not exposed to the field. A specialist recruiter in analytics stated that he rarely saw candidates with *OR* backgrounds.

> **Analytics Recruitment Consultant (Larger):** "I think probably in the last five years, obviously with the credit crunch and everything going on, Government was the biggest employer for operational research so we haven't really seen a lot of that. However, you get some really good people in it. You get loads of people that have come out of Cambridge and Oxford who go into the public sector, Civil Servant or whatever, and actually are really, really clever and really talented and get paid far less for operational research than they could working for a bank or a big retailer [...] But it's good that [the Government is] getting them, they're getting really good people."

Data mining was similarly regarded by many to be an important element of analytics. In particular, in its association with big data, it is an element that offers its approaches considerable power.

> **Software Vendor (Data Management):** "Over particularly the last 10 to 20 years the processes around data warehouse now support key operations and therefore they've become robust, and that's one of the reasons that we're saying that this discovery kind of thing isn't happening in the data warehouse [...] It's about saying, 'what are we doing at the moment and what do we do next'? So [customer] churn is a common one, whatever the industry, where you're benchmarking and saying "okay, our churn rate is not as good as we would want it to be [...] how can we detect potential churners better or sooner?" and then it comes down to can you throw more data at it, are there more sources you can include, can you include more granular data, can you include more history, can you be more refined about the models you use? There's all kinds of dimensions that that can take you into."

In terms of where *data mining* is used within the organisation, for some respondents it was becoming increasingly wide spread and cross functional.

> **Analytics Manager (Online Travel):** "Data mining is de-centralised to some extent so then I have those resources that sit in, kind of, what we call the spoke tips."

Despite its prominence in certain organisations, in others it is rarely used. Many of the organisations we described as predominantly focusing on ad-hoc data seemingly were less likely to use these methods. Again, this is probably a natural extension of the data available; data mining can only really have use when there is enough information to mine through, and as the previous example shows, one that really comes to the fore if you are utilising big data volumes and unstructured sources. However, in these scenarios the approach can create significant value, value which would otherwise remain undetected, as this case study demonstrates:

> **Software Vendor (Data Management):** "This is a real example where they looked and they found that [...] there was a general problem of [devices] that had issues with software. They detected that users kept restarting their machine. So, you find there are some issues, you drill in and see "what are the recurring patterns?" And then you discover that you can see patterns of users, three times in a row, restarting their machine. You don't have a business problem to say, 'let's find out how many times people reboot their machine' but you find that and then you discover that there's a common pattern [...] A certain manufacturer of box, a certain version of the box, certain version of the software, has got a problem. You find that you can push out a software upgrade to the customer. So again, you're tying it back to a business problem. 'If we push this out we get lower calls coming into the call centre, we get better satisfaction, we get fewer people defecting because they're p***ed off with the box'. And that to me is a perfect scenario where you say, 'we know we've got some issues, can we use this new data to solve one or more of these issues?"

Machine learning, from which *data mining* arguably represents a subset (unsupervised learning), perhaps surprisingly given the amount of attention it currently receives, was far less prominent in the organisations focused on in this study. However, for two of the companies who were using it, they were doing so extensively and considered it a critical part of their analytics functions. Indeed, for one such business, a media company, this has become not only the direction for new analytics projects, but also, they sought "to transform some of our [existing] models into the machine learning side of things."

In respect to the skills required for both *machine learning* and *data mining*, those which do not apply as readily to other forms of analytics, interviewees pointed to the increased role of computing. For candidates working for the media company, it's essential that "they've got the computational side of it [and] they've got a mathematical side of it."

The final sub-category from figure 31, *statistics*, was the area with the most widespread support across all interviews. Indeed, these were generally considered to be critical skills for all positions in analytics.

> **Government Analytics Manager :** "[Having] a familiarity with basic statistics, both descriptive statistics and sort of being able to run t-tests and non-parametric tests and all that sort of stuff. So, an understanding of statistical significance and all that sort of stuff is a requirement. Very often we're trying to judge whether things are true, or testing hypotheses on the basis of data, which is imperfect data sets, so the only way that you can do that is through statistical methods."

Two of the interviewees pointed to the growing use of design of experiments (DOE), particularly when combined with the ability to carry out testing online (for instance, in digital analytics):

> **Analytics Manager (Retail Travel) :** "So, I think experimental design I think is growing, so on the web I think particularly and our ability to do different things for different customers is increasing. One of the things I've been pushing here is ... which isn't something that the business was naturally doing ... is, you know, let's actually run an extra run and test whether this so-called improvement is actually going to [work]. So, for example, we changed our prices on [product range] recently and ... so not a decision I was wholeheartedly in favour of and ... so what we should have done is, you know, split the change, so we change it in these weeks or in these months or in these geographies and these alone and, you know, get a control set and a real set and measure what the impact is but, you know. [...] And then we looked round in sort of ... after a couple of weeks. "Actually, it looks like it's losing its money", but because we haven't got a control set [...] what we didn't know is, you know, should we have gone half-way or what should we have done. There's a discipline about running experiments, which I think is on the increase."

In summary, echoing the divide between the different types of courses presented in chapter five, generally there seems to be some difference in approach between the organisations who operate in data rich settings and those who don't. In the case of the former there is clear evidence of the growing importance of *data mining* and *machine learning*, and, therefore, an emphasis on computing skills; whereas model building and *OR* play a stronger role in the latter. However, for most respondents there was a feeling that for candidates with a sufficient quantitative background, a knowledge of *statistics*, and the ability to frame problems and understand the business context, it is relatively trivial to pick up new techniques and approaches.

6.3.3 Programming and Software

The next area of concern, most closely related to the technologies section of the taxonomy presented in figure 14, section 2.3. As with previous nodes, this topic also had multiple second-order nodes, shown in figure 32.

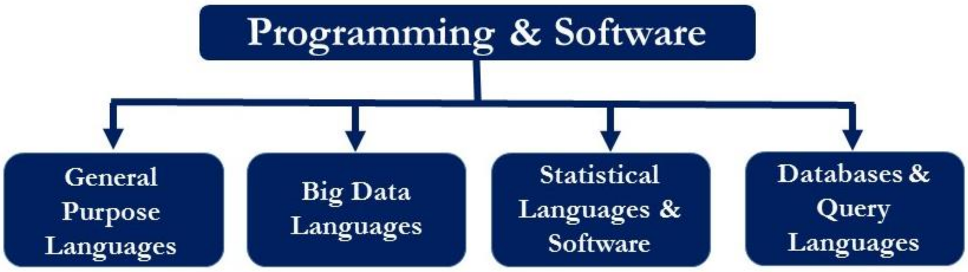


Figure 32 'Programming and software' code hierarchy

The first sub-category details *general purpose languages*; that is those languages which can be used for a variety of functions, not just for analytical tasks. By default, this is a relatively long list and includes Java, the C family, Python, VBA and several others. Of these Java was frequently cited, however this was not always necessarily regarding the analytical aspect of a project:

> **Software Vendor (Data Management) :** "There are a lot of people out there who are Java codists, who are good at Java coding, most of those people tend to have an application background and not an analytical background [...] You can try and convert those people but from an effort point of view it seems to be easier to take good analytical people who are good at SQL [and] say "let's extend what they do [using tools such as Hive and Spark SQL] and give them the capability."

Indeed, a general view was that *statistical* and *query languages* (e.g. SQL) may be more likely the domain of the analytics specialists, and that general-purpose languages were more likely to be used when these models are operationalised into the businesses processes. This was the process that was stated as standard operating procedure in four of the organisations. In one of these interviews, with an Analytics Consultant in finance, a model of this was sketched out, and an adapted form of which is shown in figure 33.

The most cited overall of these languages was VBA. Whilst the language would seem to have less 'buzz' associated with it in comparison to Python, R and other bespoke *big data languages*, there are, however, two significant drivers for this suggested in the interviews. Firstly, this is seemingly the language which most of the interviewers had the most experience with. The majority of respondents who were involved with performing analytics stated that they have used the language before, and often on a regular basis. Secondly, many interviewees stated that this language was particularly appropriate because of its link to Excel. This connection is important in that many different clients (internal and external) would provide data in Excel format; it allows the analyst to build a model which can be left with the client to use for future decision making; and finally, its visual display is one that many clients are familiar with, meaning outputs can be easily understood.

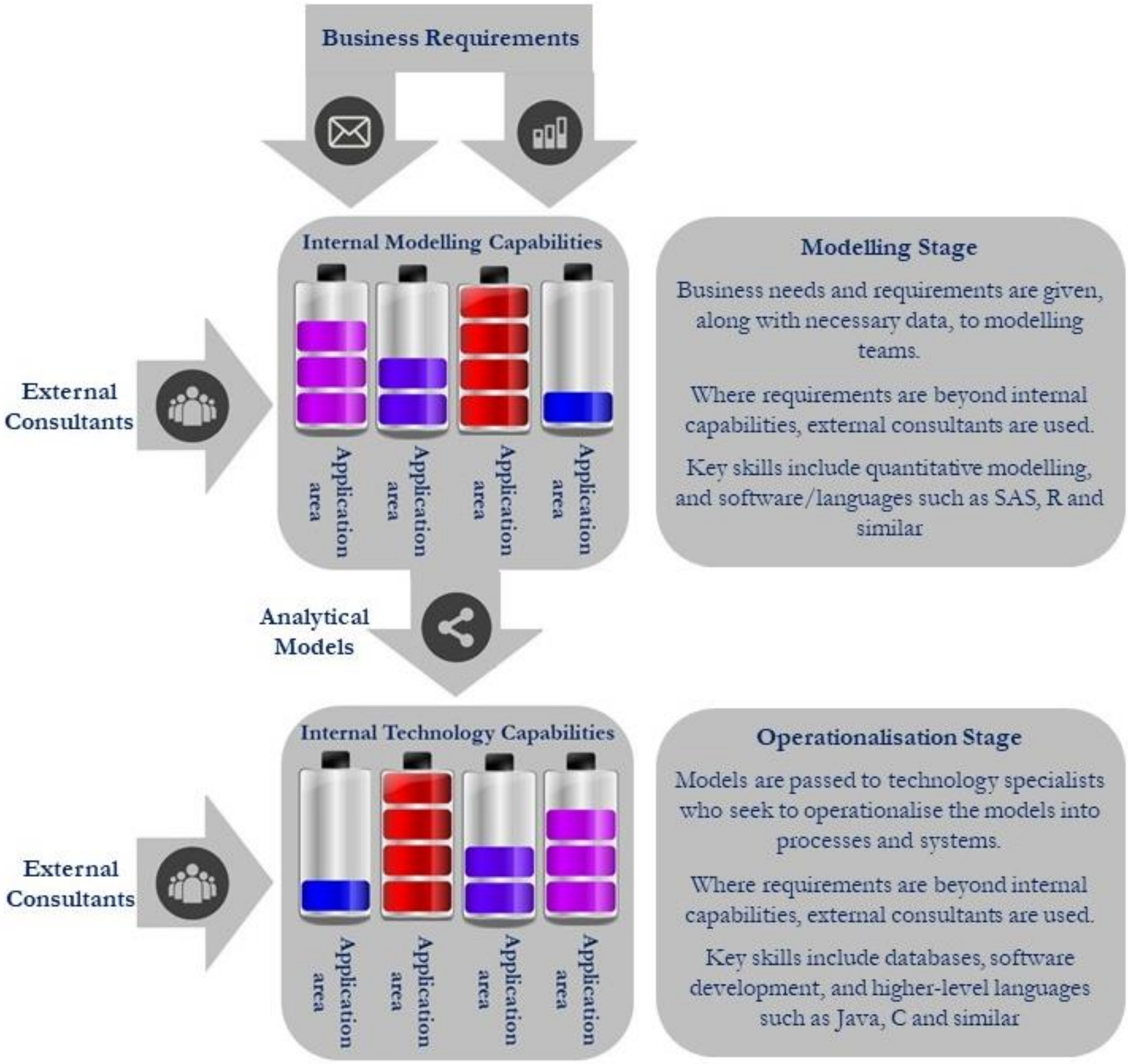


Figure 33 The use of different languages and tools in the operationalisation of analytics

The second sub-code, *big data languages*, was one that was rarely used by the subjects of this research. Indeed, most of the respondents who were using big data, instead were using it with more traditional languages such as Java and through projects such as Apache Mahout. However, several of them highlighted the growth of SQL-like languages to use with NoSQL databases, and interfaces to these databases provided in software such as SAS and R.

> **Software Vendor (Analytics General):** "People spend years building their data warehouses in something like Oracle or Teradata or DB2 or whatever they're built in. And they still exist. And they can exist for a long time. But they're going to coincide alongside some sort of Hadoop cluster. And its horses for courses. If this bit of information is going to be used forever, then put it in your warehouse. If it's transitory [use it] and throw it away next week. So, these two worlds have got to co-exist with each other and SQL clearly still plays a big part there [...] Integrate with other stuff rather than make our customers learn a whole new language."

> **Software Vendor (Data Management) :** "We see, particularly in the UK market, a shortage […] of people able to code in MapReduce. So, what we've done is we've taken […] about 70 common MapReduce functions and effectively put a SQL front end on it."

Of all the elements of figure 32, *statistical languages and software* were by far the most widely used. Whilst the R software has been widely lauded online as a language most associated with analytics, SAS was the language that most interviewees reported using. This seems particularly true of companies that deploy significant customer analytics teams, and of Government statisticians. Equally it was regarded as the language that had the highest earning potential:

> **Analytics Recruitment Consultant (Larger) :** "SAS is still the biggest technology that we're involved in. 50% of the jobs we place, normally the person has a skill in SAS, whether it's an analytical skill or it's some sort of back-end tool that they've got. [...] That's just because it's the most powerful and useful tool across the industry. It's expensive but it's really good. There are other tools but nothing is really as valuable. If someone has got SAS on their CV and they're good at it then they're definitely much more valuable than someone perhaps with R or SQL or Microsoft technology. The idea is it's easier for us to get people with Microsoft technology than it is with SAS technology."

The second most cited was the R language, and, in particular, for use in machine learning. For the companies for whom these techniques were most widely used, R was listed as the most prominent language for this. Further, many other interviewees expressed an interest in building capabilities with the language, on both a personal level and within their business function. Some mention was also given to SPSS, but with less frequency than the other two, and other bespoke software, such as AIMMS (https://aimms.com/) and Simul8 (http://www.simul8.com/).

Query languages, were also widely used. Some discussion has been given already to a perception of an increasing role for databases in analytics, however the prominence of its use was such as to warrant some further focus. Several interviewees specifically state an understanding of the language as a pre-requisite, and others that it was language frequently used in their operations.

> **Software Vendor (Simulation - All):** "Even though we don't necessarily use a live link to databases all the time, it's good to be able to get the stuff out of the database if you need to. We also actually use it for ... one particular project comes to mind where we had to analyse individual people over six properties and each property could have lots of indicators and we're now trying to slice up data and subsequently we can analyse and quantify the number of people in each of the categories and each combination of category. Well trying to do that in a two-dimensional kind of programme like Excel will drive you absolutely mental so you need to use SQL."

> **Marketing Analytics Consultant:** "Because there's so much demand in the marketplace the people with SQL skills you have to offer them a 20% premium."

Another reason given for why SQL has such importance, is not only in its direct application as a language. Many cited that through building a familiarity with the code, an individual also builds an understanding of databases and data structures:

> **Analytics Manager (Telecoms):** "If you're an analyst and you understand some of the database and SQL and stuff, then you have a better understanding of what you're receiving. Quite often you might spot where there's an issue or something with it."

Finally, there were other perspectives presented regarding the use of software and programming languages. Mention was given to many of the general Microsoft Office packages widely used in all businesses (particularly PowerPoint) and to data visualisation software such as Tableau and QlikView. These will be discussed in greater detail later in this chapter (section 6.3.7). Secondly, many argued that it was not perhaps necessary for candidates to be fully proficient in any specific tool or language, but to have a more general understanding:

> **Analytics Manager (Retail Travel):** "I didn't come out of [university] with SQL [...] I've never had any SQL training. I've just picked it up, but because I had that knowledge of what coding was about [...] Actually once you know roughly what the basis of programming is [...] you can pick up any language really."

Further some argued any significant training was unnecessary:

> **OR & Analytics Consultant:** "You just need the right sort of brain I suppose, that you can read the manual and think, "I'm looking for a way to do this and this is how I do it in C#" so then asking the right questions."

> **Analytics Manager (Consultancy):** "I'm looking for three things. I'm looking for technical modelling skills, I'm looking for ability to work with people, consulting skills, and I'm looking for business problem solving [...] And you see, no, software tools isn't one of them because I can teach that."

However, this perspective is countered elsewhere by other respondents who view programming and IT skills to be a very significant element, and one where a specific shortage has been observed:

> **Marketing Analytics Consultant:** "The only thing that we are missing normally from UK students is programming skills."

Analytics Consultant (Finance) : "Without a solid IT, analytical, technical kind of foundation no-one would stand a chance out there. When you were working with Excel and a limited amount of data I think you could actually get away with it, but at the moment you just can't [...] Being able to carry out certain types of analysis by yourself, get the data yourself, manipulate it as you wish, so that you can carry out the analysis in a more complex way and in a more agile kind of environment, that's absolutely essential."

In summary, as with data management, programming is an area where notable differences can be seen between responses. Of the languages that are most used, for the majority of analytical roles *statistical software* (such as SAS) and VBA are most highly prized, whereas for the more machine learning orientated roles R and other *general purpose languages* such as Java are of prominence. Across both categories though the most significant demand seems to be for SQL and *query languages*, both for the usefulness of the language in day-to-day tasks, and for the familiarity it builds in databases and data structures.

6.3.4 Soft Skills

The final group of skills that can be traced back to the taxonomy of figure 14 (section 2.3) relate to supporting decision making. *Soft skills* were widely seen as of high importance in the majority of organisations. As before, the topic was divided into further sub-codes, as shown in figure 34.

The first of these, *project scoping*, is essentially intended to cover all activities an analyst may engage in to capture the relevant information required to build an analytical model or to identify meaningful insights that can be deployed by the organisation. One area within this is problem structuring, and methods such as requirements gathering which are designed to capture the specifics of a client's particular issues or goals. This was held to be key in consultancy roles, where the engagement is likely to be entirely based around solving a specific problem, but also for many direct employers.

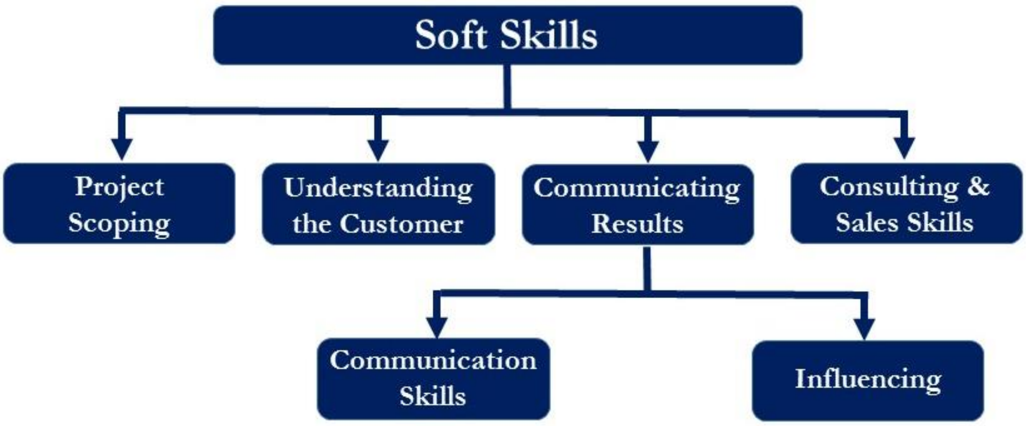


Figure 34 'Soft skills' code hierarchy

Healthcare Analytics Consultant: "The hardest [part] is finding out what the client wants. Problem structuring is ... if that's wrong at the start of the project, you might as well light a fire. And probably about 80% of the learning clients get is probably around thinking and by us developing the model. Clients think that they know what they want, but they actually don't. So, you have to have a very questioning. Always asking questions is something that I learned very quickly".

Analytics Manager (Energy): You get a piece of policy land on your desk [...] How do you unpick that? How do you turn that into something that you can operationalise? That does require analytical skills to be able to do that effectively."

Whilst such skills were deemed to be critical, most respondents didn't consider that a specific formal process needed to be followed (such as the soft systems methodology). However, one respondent stated that he did encourage its use and that he had employees with this background. Irrespectively most highlighted the importance of challenging the client and developing a fuller picture of the situation:

OR & Analytics Consultant: "We come from an ethos of absolutely not the client is always right and we'll do something that we think is wrong. We will absolutely try to sort of steer the subject. But on the other hand, the client is always listened to [...] [An IT associate will] just keep saying to me, "Well what, you know, well, just tell me what the client wants and I'll do it" And I think, well it's not that easy with that client [...] Let's think about the problem. Let's try and think for them, work with them to evolve what's to be done. He's used to very tightly structured programmer specs, very IT kind of strengths".

The second issue is being able to build an *understanding of the customer* and their requirements. This is considered to differ from *project scoping* in that this doesn't necessarily relate to formalised methods, but moreover the ability to build a fuller understanding of the client and build relationships with them. Again, this was equally perceived to be critical to delivering successful analytics projects.

Analytics Manager (Retail Travel): "There's no sense by which we sit in a back office. We sit with our clients [...] What often comes up is actually we thought this might work or we thought it works like this, but actually we haven't thought about anything else going around it [...] That approach leads to failure."

The skills involved in this though are acknowledged to be hard to teach, although possible to learn, and, as with the very nature of decision making as a category, relatively interdisciplinary. Most significantly they require the ability to empathise with the client and to appreciate their potential differences.

> **Government Analytics Manager:** “One of the most powerful things to me was when somebody did a psychometric profile [...] they were amazed to see that there were lots of other people in the world who had different psychometric make-ups. And that enabled them to think “when I’m talking, or when I’m speaking, or when I’m writing, there are other people out there who think about things in a different way to me.”

The natural counterpart to these skills are those required after the analysis is completed, the third sub-code of *communicating results*. As most analytical models are designed to make changes to an organisation, and to prompt specific actions, *communications skills* were considered crucial:

> **Digital Analytics Consultant:** “Well I think everybody needs [soft skills]. You can produce the most wonderful report in the world, and unless you can persuade someone in the company to take action on it has zero value. In fact, I even say to people it has less value. You’re wasting somebody’s time. They could be doing something more productive.”

In many cases a specific challenge for analytics recruits is to explain relatively complex analyses to non-technical audiences in a concise and comprehensible fashion:

> **Analytics Manager (Energy):** “What you’re looking for is someone who can do that translation [from analytical to business terms]. The good people within our business are the ones that understand how to analyse and how to come to the answers and the options and the scenarios, and it’s important that they then know how to present that to people who aren’t expert analysts, and are actually trying to make business decisions based on the output of that analytics.”

Particularly important to this end are general presentation skills:

> **Analytics Manager (Online Travel):** “It’s all about credibility, confidence and performance. If you walk into a meeting and jumble your words, the slides are a bit of a mess, no one can really figure out what you’re talking about, you’re not going to get the outcome you want [...] There’s plenty of examples in history where idea have been lost because a person couldn’t pitch them properly [...] It’s being able to be confident in how you communicate your ideas. Not arrogant, confident. So, all those communication skills, how to present yourself in meetings, how to cope with questions and all that kind of thing, yes is taught on the job, without a question, but if they can come with those skills they will beat their competitors.”

Indeed, the ability to effectively communicate is one area where the standard of graduates was questioned. One respondent, recounting the performance of a recent intern, stated:

> **Analytics Manager (Telecoms):** "Things like team-working would be really handy [skills]. I know they do the occasional project [in universities], which means some team-working. But that's the way it is, I mean this year we had a student and she was bloody difficult to work with frankly. Just interpersonally. I just think that would have really helped her if she'd had to learn how to work in teams, and [collaborate] and accept other people's point of views."

Beyond merely being able to communicate effectively, many interviewees cited the importance of *influencing* skills that can enable the analyst to impact changes on the organisation.

> **Analytics Manager (Online Travel):** "[An important skill is] how to influence and recommend [...] you need people that can convince, influence sometimes sceptical, sometimes not sceptical, but certainly less technical audiences."

> **Analytics Recruitment Consultant (Larger):** "[They] don't just want clever people, they want clever people that can communicate the results in a way that's going to change businesses. That's one of the hardest things to find."

However, as acknowledged by a respondent with experience in both utilities and consumer domains, the importance of *influencing* skills may vary by industry and types of projects involved:

> **Analytics Manger (Utilities):** "[In credit] I did quite a lot of stand-up presentations, 'this has been the success of our spring campaign' or 'this is how we think we should go about increasing our air miles penetration' [...] Now it's much more "Here is two or three graphs, they tell you what you need to know". That's the difference between a marketing company and a utilities company. There is not much big strategic stuff that we get to influence. We get to influence the little stuff, who are we going to chase for money next week, but not how are you going to do it or why are you going to do it."

The final aspect, leading on from the previous, was the requirement for *consulting and sales skills*. The argument made in several of the interviews was that these skills were growing in importance, and indeed in some cases becoming essential to the effectiveness of an analyst:

> **Analytics Recruitment Consultant (Larger):** "Gone are the days when you had someone who was a salesy, articulate guy in one department and they did all the selling in the business, and you've got the geeks who are doing maths somewhere. Those geeks now have to be as good as those guys at selling and changing the business as well. They're not necessarily going to be at people's desks driving through a change programme, but they've got to really prove and sell to the Board why this data that they've found can make a massive difference and ultimately save the company money or make the company money."

One respondent, a consultant and software vendor, describes how he would look for candidates who have worked directly in customer facing roles in restaurants or shops. Whilst such experience is rarely discussed in the analytics literature, the argument is it is this type of client-facing, service orientated work that both gives the candidate confidence to manage interactions, but also key skills such as upselling, handling objections and identifying the benefits of a product (whether that be a bottle of wine or a complex analysis).

Consultancy skills, those beyond the aspects already discussed in this section, were also deemed to be of importance, and not just for consultancy positions. The nature of the field means analysts are often working on a variety of projects, and often for different internal or external clients.

> **Analytics Recruitment Consultant (Niche):** "[We look for] some of the consultancy skills. Consultants are sometimes ... an analogy is those who can turn ... can investigate problems and opportunities and turn those into recommendations and decision support. You need that set and that's your typical management consultants and for your analyst who [succeed in] businesses."

> **Media Company Analyst:** "In consulting, you would go into different organisations, you'd identify opportunities or maybe they've already identified opportunities where data can support them but they might not have the skills in house to do it, so you'd be going in and conducting a project using data to provide some strategic direction, operational direction, give them some insights, which could lead to them doing some more work in another type of area. This isn't too different [...] our work is forward facing. We design projects that look at the way we're working now and provide insight on what works, what isn't working, where we could do things differently, test new approaches and provide insight [...] If we're doing something that's not going to result in potentially doing something different then it's not for us."

In summary, whist the degree to which *soft skills* are employed in analytics positions does naturally vary, all interviewees highlighted *some* importance, particularly in moving beyond entry level positions. Aside from the very large employers, many highlighted that they found this an area that they struggle to find as easily as they do more technical skills (as the larger companies effectively have the 'pick of the bunch'). Furthermore, there are suggestions that demands may be rising in this regard, and that communication skills alone may not be enough, with employers increasingly requiring *sales skills*, *influencing* skills and the ability to drive change in an organisation.

6.3.5 Other Skills

The final grouping of 'skills' codes, essentially a catch-all, details the other requirements given by employers. Three major areas were particularly frequent in the interview discussions, as detailed in figure 35.

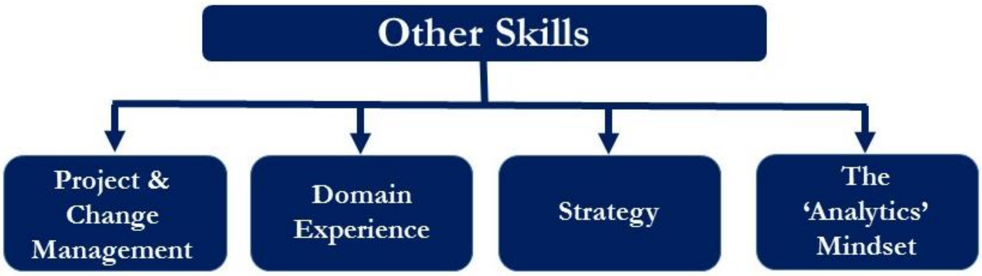


Figure 35 'Other skills' code hierarchy

From the first of these codes, *project management*, was regarded as important in virtually all cases, but variation was observed in the extent to which these processes were formalised or ad-hoc. For some respondents, it was more important to have an overall appreciation of what *project management* was about than it was to have certification or training in one approach or another.

> **Government Analytics Manager:** "Being able to do some simple project management tasks is quite helpful, to understand that projects have a scope and a beginning and a middle and an end. Most of our projects you don't need to have huge Gantt charts [...] but you do need to understand that there's a list of tasks and schedule those tasks."

> **Analytics Recruitment Consultant (Larger):** "Very rarely would someone say you have to have a PRINCE2 qualification or something like that [in analytics roles]. Within IT you have to have a formal qualification, don't you, and methodology and all this sort of stuff ... Whereas most of the analytics projects we see, you just have to have had some experience of project management, and they never really quote to us "you have to have this qualification, you have to have a training course in this methodology". I'm not saying that won't become more formalised – it probably will – but at the moment it's not."

However, for others such skills were more important, and not just for those involved in software development. Indeed, for one respondent, involved in analytics software and consultancy projects, he regarded such skills as one of the three "pillars" of skills that he required of candidates, alongside technical skills (both analytical and programming) and commercial skills. The approach they used was:

> **Software Vendor (Simulation - All):** "PRINCE2 form but basically a light version of PRINCE2. What we use is Microsoft Projects ... project plan, fits on Milestones, scope out the steps that we need to follow, who is responsible for what, plan and map your risks, use

> it to track progress and use the reporting functions to produce charts and graphs and send it on to the customer [...] Basically, what I need to know is I need to see that my consultant is able to demonstrate that we have the plan over 25 days. If they come back with the project for 40 days, there are some things seriously wrong."

The most commonly cited methodology was Agile, or adapted versions of Agile:

> **Media Company Analyst**: "[For] the analysts there's a more agile methodology as in they've got projects in the pipeline and we have a weekly meeting where we say right, this is what's in sprint, this is what's upcoming, how are we in terms of resources, where do we move things to, where might we need to push things out."

Additionally, some saw demand for *change management* skills, or at least awareness of these issues:

> **Analytics Manager (Utilities):** "[Its] about being able to demonstrate we're about transforming business processes rather than just coming up with interesting findings. To get transformed into real tangible financial benefits [...] Things like benefits management, change and business transformation sort of skills could be good skills to have from universities."

> **Analytics Manager (Retail Travel):** "We need to recognise that that's what we're doing and we need to recognise the importance of bringing people on the journey and that's what ... and the importance of managing that business change."

The second area of interest, *domain experience*, was again something regarded as of high importance for many interviewees. As such, several organisations sought to expose new recruits (particularly those at a graduate level) to different areas of the business, even those that were not regularly employing analytical methods. However, it was generally felt that such experience is one that fundamentally needed to be acquired over time, and not necessarily something easily taught:

> **Analytics Recruitment Consultant (Larger):** "You don't just need analytical [skills] you need to understand all the regulations in that industry, how the business works, all the types of techniques and modelling which is corresponding. So, you might be data-savvy and very good statistically but actually picking up all the industry information can take years."

The importance of this experience, however, does seem to vary across industry types. In particular, respondents working in healthcare and utilities highlighted its importance, whereas for other industries, and for many of the consultants (who by default will often work in multiple sectors), it was of lesser significance. Fundamentally though such experience can have additional benefits, in that it makes problem structuring and effective communication within the organisation easier to achieve.

Additionally, some highlighted the growing importance of *strategy* and strategic thinking, both in respect to the organisation as a whole, and to how analytics is structured within it. For some this took the form of being able to identify which products were likely to bring the most value, and 'weed out' those that were less likely to generate significant benefits. Additionally, however, a need was highlighted in better defining how analytics and information can be used to support and to transform the business:

> **Technology Consultant:** "It's about using the results of analytics that should be focused on [...] how you should use technology to enable business strategy. That's what business schools should be teaching in the IT modules. How do you use information and information technology to transform a business? Either the performance of the business or the structure of the business, but business transformation ... and analytics is a key part of that. Right, so for me there's two elements to the question: so yes, you need to be teaching people to do analytics and to develop analytical apps, which is critical, but it also the MBAs which is exploiting information and using information for transforming business."

Beyond more general business strategies, the ability to be able to structure how analytics was managed in an organisation was identified as an important and emergent demand, and one to which analytics specialists themselves may be best placed to enact:

> **Analytics Consultant (Smaller Management Consultancy):** "If you trained analysts to be really up to date with the latest technologies and techniques but to be able to structure and shape analytics in organisations, they will become the most employable people, because that's the problem."

The final skills that were most in demand amongst interviewees, can be loosely categorised under the heading of *the analytics mindset*. One aspect of this is problem solving and critical thinking skills:

> **Analytics Manager (Consultancy):** "The other key thing is problem solving. I do think that is the role of a university to teach because we are problem solvers and one of the things we test at recruitment is can people do problem solving, and good courses in universities teach people how to problem solve."

> **Analytics Manager (Public):** "[Analysts will] never be given a very clear brief on what you need to deliver, it's much more kind of critically thinking [...] It's not an easy thing to train [...] For our junior level analysts, who will be fresh from university or may have two or three years' experience, you can teach them to program, you can teach them new techniques and so on, but you can't teach them to structure problems and think independently."

Another element is an understanding of modelling and the relationship between analytics models and their parameters, and the reality of an organisations. For one interviewee, discussing a staff member working at a client company, such a way of thinking can exist even without an analytics background *per se*:

> **OR & Analytics Consultant:** "[The staff member is] really good and savvy and I would now say is a good ... just a good modeller. She couldn't have done what we did in the coding. But she gets it and she gets that a model's a model. She's really quite junior, you wouldn't recognise her. But somebody that's been with [the company] for 10 years might be way down, might have lots of context and domain knowledge, but be pretty poor as a descendant in terms of if they inherited her [work]."

This idea, that fundamentally there may be some aspects of analytics that a more innate than taught, was echoed elsewhere:

> **Analytics Manager (Utilities):** "It is that nature/nurture thing isn't it, can you teach people to think in the way that ... I'll tell you a story because it's funny and it kind of illustrates a point. My dad is an accountant and my mum is a maths teacher. So, me and my brother and my sister all did maths degrees, my middle brother is a lawyer. So, we're at the kitchen table one day, I'm aged about 13 or 14 and [my brother] is 8 at the time and we're discussing how much milk a corn flake can absorb, and how you would measure that, and [my brother] just turned to mum and said, 'Mum am I adopted?' [...] I think you can teach quite a lot of it, particularly the stuff about understanding what the problem is, if you're going to think about a problem structure, I think that can be taught, I think presentation skills can be taught [...] But I suspect that you can't teach people the innate ability to be able to understand the multi dimension analysis, you can teach people how to do it, but if they haven't got that way of thinking then it's probably never going to be there."

Overall, as with many of the other skills in this section, there is evidence of increasing demands, and for skills that would not necessarily be associated with traditional OR teaching. Equally, there has been some questions raised as to the extent that all the necessary skills *can* be taught, of if some are more innate. Such questions provide a natural progression into the next topic of the template: education and training.

6.3.6 Education

As the natural counterpart to discussions on skills and skills shortages, most respondents also had opinions and recommendations on how we educate and train potential recruits in universities. Again, this node also subsumed several child nodes, which are presented in figure 36.

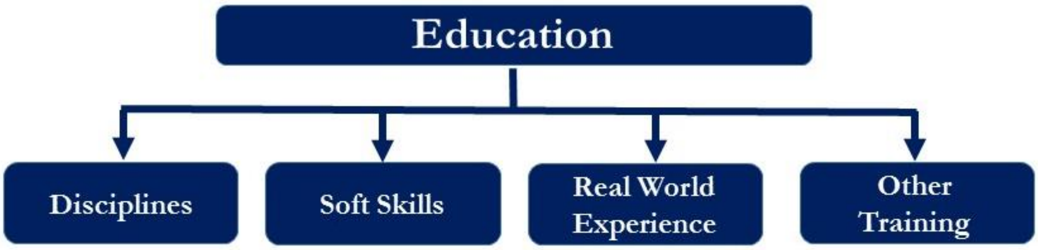


Figure 36 'Education' code hierarchy

The first of these concerned the *disciplines* from which recruits were drawn. Many were generally open to a variety of disciplines, possibly reflecting the relative lack of candidates in the market and/or the relatively interdisciplinary nature of analytics (as discussed in section 1.2). For most the main criteria was for a "quantitative" discipline. Obviously, this would include the more overtly so (such as statistics or OR), other disciplines that have a clear connection to quantitative methods (such as computer science or physical sciences), and others that were perhaps a little less obvious connected. Examples given of the last type included geography, music and philosophy. Many of the respondents ideally looked for a master's degree. As discussed earlier, for one company the ideal was for an OR masters, one interviewee discussed the benefits of a current MSc in Data Science, and for another, who were specialising in machine learning, the preference was for advanced computer science degrees.

Although many looked for such specialisation, for a few respondents having a too specialised degree, conversely, could be potentially problematic:

> **Digital Analytics Consultant:** "My first company didn't like computer science graduates because they'd always had perfect stuff. So, here's the spec change when you're writing a program ... Because it makes you think in a totally different way I wouldn't have done it like that if I realised there was going to be a spec change. I wouldn't have tried to be so clever because in fact all my clever stuff all my complicated code I can't unravel it."

Whilst this example specifically details programming and computer science degrees, a similar danger can be extrapolated from this, when analysis based degrees use artificial datasets, something which will be discussed later.

For some interviewees, the major concern was pre-university, specifically A-level mathematics:

> **Analytics Recruitment Consultant (Larger):** "A lot of the big recruiters in our area, they look at their Maths A-level as much as they look at their degree, and they want straight A's at A Level really, in numerate disciplines."

> **Analytics Manager (Consultancy):** "What's making my life harder is there's not enough people going in the funnel at the top. There are not enough people in this country doing maths A-level."

Such concerns may well be valid, and indeed a potential contributory factor in any skills shortage in analytics. However, they are slightly outside the scope of this research, but an area that may be worth further exploration in future research.

The second area of concern was the teaching of *soft skills*, as they had been held to be of high importance for employers, and not as easy to teach or test for, or to recruit for:

> **Healthcare Analytics Consultant:** "It's difficult. I think soft skills is very difficult [...] You learn them, but you need to learn by doing them, not trying to teach them."

> **Analytics Recruitment Consultant (Larger):** "There's no point searching for people based on soft skills, you can only search people based on objective skills, because you've got a job spec and classically you might see the few things are "outgoing personality, great client-facing skills, great team player". Now how many people if you asked if you're outgoing, how many people would say "no"? How many people would say "no, I'm not a great team player". Everyone is, or everyone thinks they are."

As discussed previously, for some this was an area that they felt universities could improve on:

> **Analytics Manager (Online Travel):** "That is something that I've never found yet in a university course. So, when I hire junior people, hire an intern for example, who is now at a full job, that's what you have to teach. So, I find the courses I run here, a large proportion of them are more the softer skills [...] When the candidates green arrives through the door, yes, they've got some good technical, probably even better technical skills than people [already in the job], you know, they know the latest things. Universities do generally a good job at keeping up to date with the latest techniques, but often the challenge that those techniques, the applicability of them can be challenging, but frankly the people don't have the skills to be able make use of them and have them and you have to then do a large job at making them useful."

Two approaches that some interviewees thought can help build these skills was through project work (in teams) and through delivering presentations:

> **Analytics Manager (Telecoms) :** "[Lecturers] are more attracted to the technical stuff. And I'm not suggestion you'd necessarily try to rate how good someone's team-working skills was, but just by the very nature of giving people more team-based projects they're going to pick it up."

Healthcare Analytics Consultant: "[In my degree] every week you [...] were up presenting [...] When you're doing it every week, you eventually get over your nerves [...] For me, that was one of the most beneficial things."

The third area deemed to be of importance was to give students greater *real world experience*. This was indicated in the earlier discussion of programming and computer science degrees, but equally was something many interviewees considered was an area for potential improvement:

OR & Analytics Consultant: "We really need people to operate in the real world [...] I think [that is] what we have found [to be] lacking [...] You come from a university learning environment where it's like "here's the problem" and [they] spell it out for you."

Firstly, many highlighted the use of artificial datasets as a potential problem, datasets which had been designed to demonstrate a specific problem and arranged accordingly.

Analytics Manager (Public): "One of the things that we have to teach them on that is working with messy data sets, because they work on very idealised data sets in their training and then when they come in and they're in a different world. You've got messy data spread across a number of systems, you've got to make some quite big assumptions around this and that kind of critical thinking. I do wonder if the universities could do more in that kind of working with rubbish data."

Analytics Manager (Health): "I think they should get an unstructured problem, I really do. My own view is that far too much analytical training is here's the technique, turn the handle. Oh, and here's the pure data source as well. You never get pure data sources."

For the reasons discussed in the earlier sections, such problems are likely magnified by the increase of messier consumer and machine generated data, often associated with internet sources. However, as acknowledged by one of the interviewees, real datasets too may present issues:

Software Vendor (Simulation - All): "It's a tough one because you don't want your students to be spending ages on cleansing data [...] They're not going to like you, they're not going to like the course. They're going to take away that this is a painful exercise to begin with, they won't see the value."

The second approach suggested in the interviews is through real world projects. Obviously, this is something many universities already offer; indeed, many interviewees were involved in student projects from various degree programs, but one many felt was potentially key in developing rounded skills. Of course, this come in a variety of forms. At the most involved end were internships and consultancy projects based within real organisations. Whilst the management of such activities presents a not insignificant workload on universities, particularly in finding suitable projects and partner companies, many respondents suggested these activities could potentially bring significant benefits:

> **Analytics Recruitment Consultant (Larger):** "I think business would love that as well. Because there's so much need for this ... three months where they can come in and do little pieces of analytics for them and report on ... that's great for them, for their course, but it's also great for a business."

> **Healthcare Analytics Consultant:** "Internships are by far the most valuable thing I did [in my degree]."

Alternatively, simply presenting real world problems and case studies in lectures were seen as something that can create a lot of value.

> **Healthcare Analytics Consultant:** "[I had] a lot of good experience [on my degree] in terms of actually working with real clients. So, they're very good at giving you live projects to work on. [...] Sometimes companies wouldn't show up or anything. They would just send "here is a question, answer it", but you'd have to deliver it as a real project."

Finally, we considered some of the alternative forms of training and courses, particularly the rise in MOOCs (Massively Open Online Courses). Some thought there was great potential value in such developments, and several respondents were taking part in them themselves, in particular in allowing potential recruits access to new learning and opportunities:

> **Analytics Manager (Online Travel):** "Otherwise the vast majority of people [in the team] have an interest and have taught themselves. There are plenty of online tools and technologies, of course. Code Academy is a good example of those tools. So, there's plenty of tools out there but people come to it with those kind of background, or have been more normal developers, standardised developers in Java and places like that, and now move across into the data space."

However, there were concerns expressed about the credibility of these options, something which may limit their usefulness as a genuine route into employment over degrees and similar education.

> **Analytics Recruitment Consultant (Larger):** "How credible [are online courses/certificates] in an organisation? If you don't know about it, it's not credible, is it?"

Overall, there were clear recommendations from the interviewees on how education and training can be improved from their perspective. These suggestions were included in the later interviews with educators and academics (chapter seven) to evaluate whether they can be employed in university teaching, and what potential barriers may exist.

6.3.7 Outputs

Another significant factor, both on how analytics is used within organisations and the skills requirements for recruits, concerns what outputs analytics teams produce. To some extent this has overlaps with previous discussion, in terms of the skills required to communicate results, but there are other possible outputs which need to be considered. Four sub-codes were identified, and are presented in figure 37.

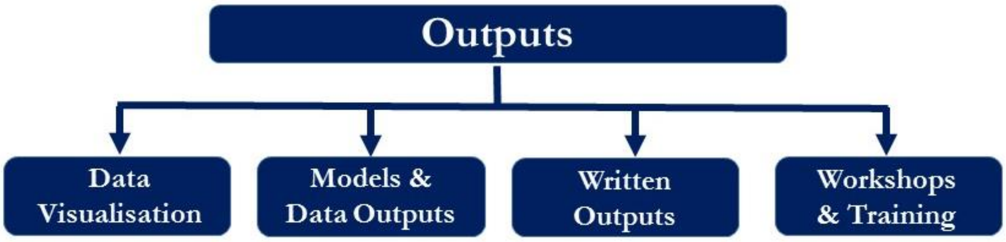


Figure 37 'Outputs' code hierarchy

The first of these regards *visualisation*. In practice, this can include a wide variety of techniques and tools, the main ones of which will be discussed. At the more 'low-tech' end this will include PowerPoint slides, charts and graphs. Despite less hype and attention, such tools were still very widely used. Equally many felt they remained very powerful and influential, if used effectively:

> **Digital Analytics Consultant:** "Often the best visualisations are the very simple ones. One of the best ones in web analytics was for [client] ... And the graph I showed them, was just time of day stats for people on their website. So, started off like that, went up to this at about 9 o'clock, stayed there until midnight. He was 80% between nine and midnight. And they said, 'oh wow, our web people have been saying turn it off at 6 o'clock because nobody looks at [the site]' [...] That's an excellent example of how a report, and it was incredibly simple, but tells you something ... Look it once, look at it again in 12 months' time to see if it's changed ... But it tells you something ... Slap your web developers on the wrist and tell them to sort out."

> **Analytics Manager (Retail Travel):** "A member of my team was using graphs and the message of the presentation stood out clearly in the graph [...] This was 100% times more powerful as a visual than it would have been just as a spreadsheet of numbers."

Additional to these more ad-hoc methods, in many cases systems and software is used to increase the visual impact of analytics. Many respondents reported the use of dashboards, however predominantly the main challenges faced here concerned selecting the right data and metrics to use rather than explicit visual concerns. Less frequently used by the interviewees, but something the majority expressed an interest in, were *data visualisation* tools such as QlikView and Tableau. For most respondents, this was an area of great interest, but not necessarily full established yet.

However, this was an area where many were seeking to employ new ideas, and new staff to help to bring them in. For one company, this had included employing a graphic designer, and another had artists working on some projects, but for most this was an additional ‘nice-to-have’ skill aside those they already looked for in recruits.

> **Government Analytics Manager (Finance):** “We will ask people to develop in those visual displays of information but whether we would desperately need them to come in ... I guess from your point of view, from delivering students or producing good quality students, then yes, that would be a big tick in the box, but whether we would reject somebody if they didn’t have that is a different question.”

Overall, most recognised the value in visual outputs, particularly as a communication tool but also to support analyses and discover new insights. However, it was suggested their work in this area was not particularly formalised, relatively early in its progression, and/or not necessarily the most important element on which they recruited.

The second output seemingly in regular use were *model and data outputs*, that is scenarios where working models and tools were left with the client for future use, or indeed the data produced by the model was left with them for further internal analyses. The former of these was a relative commonplace occurrence for many in the sample, and predominantly this was in Excel:

> **Media Company Analyst:** “If you’re going to produce some modelling that they’re going to use, you need to give them something that they can take back into the business, which sometimes limits you to Excel.”

The benefit of this approach is that the impact of the analytics can be extended beyond the lifetime of an individual project or intervention, and give the customer a tool which can bring longer term value. However, many cited the potential issues of clients misusing the tool thereafter, an issue that can only be managed through providing adequate training for the client, which again has obvious skill implications and will be discussed later in this section.

Instead of a working model, for some clients the preference was for *data outputs*. Typically, these clients will be one’s who regularly work with data, so stated examples were finance or sales teams. In many ways, such an output requires the least amount of effort for the analyst, however some identified the value in ensuring data displayed in Excel or similar had the correct visual and data layout to be understood by the client.

Thirdly many interviewees expressed the importance of *written outputs* such as reports and papers. Indeed, this was another area where issues were identified in the quality of recruits:

> **Analytics Manager (Public):** "We tend to want our analysts to be able to write well. We'll accept that the really good analysts maybe aren't as good as some of the more generalists in the office, but they do have to be able to write well and they definitely have to be able to communicate internally [...] One area we've had concern with, is probably not so much the internal communication and presenting, but it was actually just the quality of some people we've recruited, their written work, over the last few years. [Often] foreign students whose English just wasn't up to standard with writing [high-level] reports [...] It's not universal. We've got lots of foreign nationals who are extremely talented and write very, very well, but we've had certain individuals who just weren't quite there in terms of they just couldn't express their ideas clearly enough in a way that [the audience] would be able to buy into."

Furthermore, the ability to be concise was championed in multiple interviews:

> **Analytics Manager (Online Travel):** "What one of my mentors in my career made me do was if you write something in an email, can you halve it, what you wrote, to such an extent where it became almost a way of life."

Finally, some of the interviewees highlighted the importance of *workshops and training* as a tool to help implement the changes suggested in analyses. Again, this brings in a range of other skills above and beyond the soft skills presented earlier in this section.

> **Software Vendor (Analytics General):** "The big area I think that we try and get people comfortable with is facilitation of workshops and running workshops, which is quite an interesting thing for a relatively young person, because quite often you've got all these grey-haired people around you, older people in the business who know what they're doing. You're trying to elicit requirements and get understanding in a group, you've got to have confidence and be reasonably assertive, but also have a lot of empathy. You've got to be able to shift your approach depending on who's in the room and how things are going."

> **Analytics Manager (Online Travel):** "There's facilitations and they have to be able to run a meeting. [...] They have to be able to think on their feet and have structured discussions, arguments, effectively."

One respondent had developed an unorthodox approach to help train staff for this:

> **Software Vendor (Analytics General):** "We don't do a standard presentation skills course, we do a workshop with a bunch of people who used to be actors, a whole different style of things. And our questioning and listening course is with someone that used to be a police negotiator, so again giving you a whole new set of skills. How to ask a question without asking a question."

Overall, we can see a wide range of outputs required from analytics teams, and significant differences between some organisations. Many of the data-rich organisations and vendors were more likely to promote *visualisations* to display analyses; many working in Government highlighted the importance of *written materials*; and in other sectors (such as healthcare) facilitations and *workshops* were particularly important. Across almost all sectors and companies, however, the role of presentations was significant, and most saw *data visualisation* as growing in importance.

In summary, this section has demonstrated something of the interplay between skills requirements and the way that analytics functions within organisations. Clear differences have been observed between the structure of different organisation's analytics teams, as well as the outputs they are expected to deliver, there are important considerations of which skills need to be delivered to which candidates. In other words, the skills requirements for candidates will be dependent on the approach the recruiting company takes, making it far harder for educational provisions to take a 'one-size-fits-all' approach.

6.3.8 Analytics Trends

The interviewees were also asked to consider current trends in the analytics space, as they perceived them, and the areas where the interviewees forecast growth. Whilst there were many suggestions in this area, the majority could be summarised into three main topics, as shown in figure 38.

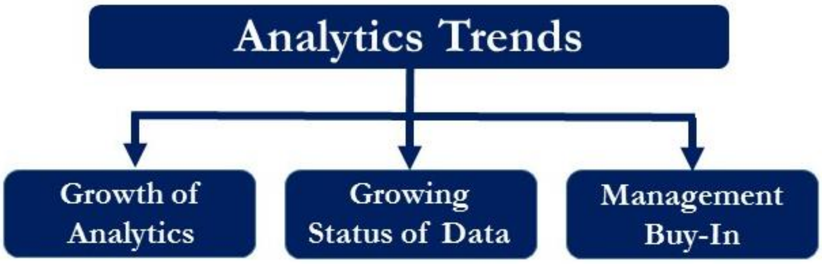


Figure 38 'Analytics trends' code hierarchy

Firstly, many respondents reported they had seen growing interest, awareness and demand for analytics in their organisations, as had been suggested in the earlier literature and analyses.

> **Analytics Recruitment Consultant (Larger):** "We've seen it obviously get busier and busier and the whole phenomena of analytics has spread across the world in different industry sectors and different applications for analytics, from credit risk and masking analytics to actuarial sciences and data science. There's so many algorithms and stats-based decisions basically that our business is in good shape and it's growing steadily."

> **OR & Analytics Consultant:** "Data, and having more of it around, is making the questions more interesting for us. It feels like, oh my goodness, if OR and analytics ever had its day, it's now. It's like it's just the time to be out there doing it, because the things that were theory and textbook and we did tiny examples in those days, you know, decades ago, you know, it's just fabulous to think we can now actually practically do them in real time and solve something, re-optimise in five minutes."

Most respondents felt that this would not simply be a fad, and that this growth was likely to continue, providing their departments could continue to prove the benefits it can bring:

> **Analytics Manager (Utilities):** If we can show that we are responsible for delivering business benefit then we'll be able to expand but if we can't it'll get chopped. [Its] about being able to demonstrate we're about transforming business processes rather than just coming up with interesting findings. To get transformed into real tangible financial benefits. That's more broader than traditional OR-type teams."

For several interviewees, a significant factor is the *growing status of data*. The internet, the argument goes, has made data and information so accessible to everyone, that it is significantly changing how we as a society operate, and has rapidly increased our expectations. Many see this as one of the drivers of analytics, and one that changes the way businesses function and may inspire more students into analytics and quantitative degrees.

> **Software Vendor (Information Technology):** "In terms of the expectation that people have about accessibility of information, the ability to Google anything. That in itself will shift the management of organisations to becoming much more information-aware than they are at the moment."

> **Analytics Manager (Utilities):** "If you type in a flight number into Google you'll see a little dashboard coming up saying when it's due and how long the flight time is and everything [...] We can teach teenagers that this data they are using, this application that they are using has got data behind it, and there is a career path in being able to present that data in an interesting way."

Whether this is the principal driver or not, several respondents considered there to be significant change in the attitude of senior managers, and significantly more *management buy-in*, a trend that was likely to provide opportunities for both individual analysts, and for analytics as a whole:

> **Government Analytics Manager (Finance):** "Internally a lot of senior people are talking about analytics. Across Government lots of people are talking about big data. My guess is that not many of them realise or understand really what it is, so there's a bit of hype there.

And I think from our point of view we need to show some good results in the next six months of what analytics could do, and I think if there are some good results then yes, it could take off quite a lot."

Analytics Manager (Utilities): "It mainly comes from our CEO, the CEO has pushed it, and I don't think it would necessarily have worked [otherwise], because it is such a new thing and requires a large level of investment, that if somebody lower down the organisation said, "oh, you know, I want to do this big, big data strategy", they wouldn't have been able to drive it forward. If the CEO wasn't picturing the same sort of thing then they would have been limited to the Excel or whatever, but because the CEO has this big, big idea of big data and what it could bring and if it can work in other industries why can't it work here, it must be able to work here."

Overall, the main trends suggested by the interviewees are of continued growth and continued opportunities for analytics. Whilst this is an obvious opportunity for OR and other related disciplines, it also gives greater incentive to ensure education and training is delivered in the best possible way to maximise these opportunities.

6.3.9 Internal Organisation

Across the different companies included in this part of the research a variety of different approaches were taken to how analytics was managed in their organisations, and the different personnel (and therefore skillsets) used in each stage. These different methods of structuring analytics and analytical teams were categorised into five distinct systems, as presented in figure 39, and each which will be discussed in sequence.

Figure 39 Different approaches to the internal organisation of analytics

6.3.9.1 The 'Devolved' Approach

The first approach we identify is one were the separate aspects of analytics were each performed by separate teams or separate individuals. Although there were some teams that displayed some semblance to this model, effectively only one respondent precisely fitted this category. Nevertheless, the approach is clearly a viable one, and was discussed in multiple interviews. For this respondent, a digital analytics consultant, his business was structured as a virtual enterprise (e.g. Davidow and Malone, 1992) whereby he, and a staff member employed, offered a "virtual web analyst" service and additional resources were used, depending on the project, on a purely ad-hoc basis and drawn from the respondent's network of fellow consultants. Each of these have their own specialisations, and therefore can bring different benefits to different projects.

The fundamental benefit of such an approach is that the teams or individuals involved could specialise in their individual element, which though it may sound trivial, in fact can make recruitment considerably easier for direct employers, and for consultants allowing for increased specialisation. On the other hand, the principal downside is the lack of integration and visibility between different team members. Such issues can negatively impact on communication due to a lack of appreciation of the working practices of other individuals or teams.

6.3.9.2 The 'Unicorn' Approach

The second approach observed, one that has resonance with the earlier discussion of the skills requirements of data scientist roles, we label as *the 'unicorn' approach*. In this model individuals are involved with all the core elements of analytics and are required to have skills in all areas; from technical to softer skills. This was the dominant model in two of the cases studied, although to some extent other cases also employed a not dissimilar model. For these two, a government data scientist and a software vendor and consultant, the individuals were involved in each of these aspects, and effectively took a project through all of these major elements in their entirety.

The most significant benefit of employing such an approach is that it removes any issues with miscommunication between spokes, and gives the individual a complete view of the project.

> **Analytics Manager (Retail Travel):** "If the analytical person takes the responsibility for the whole end to end, then you also start to spot the interaction between, yes, the process, the technology and the people […] you've got to be able to do end to end."

However, the most significant issue is the difficulty in finding individuals who are truly competent in all of these areas; areas which require wide ranging and diverse skills. Indeed, some respondents called into question the feasibility of finding such individuals, or at least those who had a *genuine* depth in each of these areas:

> **Analytics Manager (Health):** "I think managers in analytics are going to have to realise that superman doesn't exist in one person [...] the report that I was reading was that the analytics professional can do SQL but all these other techniques as well as communicate with all and sundry. That's superman. I don't know a single person with all those skills wrapped up into one."

Overall, such an approach does have clear and obvious benefits, but is one that makes the demands on a candidate's skillset all the greater. For universities and educators, the question this poses is the feasibility of covering all of these topics in a relatively short time space. If this is not possible, which skills would an individual need to acquire elsewhere, and which skills are easiest to acquire through experience and/or other forms of study?

6.3.9.3 The 'On Demand' Approach

The third structure observed, likely to be the approach most familiar to most OR analysts, is one where the main analytics team would lead the whole project, but other resources (specifically those specialising in technologies and data management) would be utilised in an ad-hoc fashion. This was the most widely used approach observed, although this may have been influenced by the relatively high proportion of respondents who had an OR connection. The necessary skillset for analytics specialists in such a system were described by one respondent, a software vendor, as "the hybrid of the [...] business and analytical person."

The strengths of such an approach is that it combines some of the benefits of the 'unicorn' approach, without the requirement of having all of the skills this entails. As technological support is provided elsewhere, such as the extraction and pre-processing of data or the coding of specific tools to be used, there is less requirement for IT and programming skills. However, it is likely that some coding skills would be required in such teams, though potentially in software such as SAS and Excel (VBA).

The major drawback of this approach however, is that the data and technological layers remain somewhat separate from the modelling and decision making layers. An argument may be made that this is something that may become increasingly more problematic. As the amount of data grows, and the complexity of managing it increases, these elements become more significant to the success and potential power of analytics projects. However, the counter point to this would be that as such layers become increasingly specialised, it may be necessary in respect to available resources and skillsets.

6.3.9.4 The 'Operationalisation' Approach

The fourth approach has many similarities with the previous, in terms of likely personnel. Again, predominantly this system is similar to most traditional OR -type projects where a modelling team, possibly with assistance from business intelligence/database teams and other technical support, manages the full process. However, the difference in this case is that the goal is not a report or recommendation, but rather the operationalisation of the model into enterprise systems. Therefore, the primary 'on demand' requirement is in terms of outputs, the incorporation of the analytics into enterprise systems. For this reason, the approach could be considered a special case of the 'on demand' approach, but the regularity with which this demand seemed to occur (i.e. that the majority of analyses were operationalised in this way), suggested a separate approach be warranted.

Essentially, this is the same process as described visually in figure 33, which also highlights some of the skills required (particularly in respect to programming languages). As such, this approach is perhaps more suited to direct employers than to consultancies, and the tasks are just as likely to represent part of continuous process improvements as to solving one-off issues.

Several organisations were employing a system similar to this, two of which appear to use both this and the 'on demand' approach depending on the task in hand. In terms of the advantages and disadvantages this approach has, in general these are the same as in the previous section. However, there is an additional skills requirement in that the results, and indeed the models themselves, need also to be communicated to the team responsible for operationalising them into the company systems:

> **Analytics Manager (Retail Travel):** "The more that OR moves into, or one element of OR moves into, developing decision making technologies that sit at the heart or part of an application, the more important the ability to write a good set of requirements becomes and to think about what needs to happen to make it more robust and that almost goes back into in terms of, if you can formulate that solution to those requirements, if it's something that can be easily coded or you know roughly what it's going to look like, you shorten that large circle even more because you can have a sensible conversation as well with a developer on the other side."

6.3.9.5 The 'Technical + Business' Approach

The final approach observed, particularly prominently in two of the cases but with elements apparent in others, is to assign the technical responsibilities (including data, technologies and quantitative methods) into one team/individual and the business-facing side to another. The companies most obviously employing this model were in marketing analytics and media respectively. In the case of the former, the business-orientated employees were primarily sales-type people with little analytical skills (certainly as a prerequisite though undoubtedly some awareness would be developed with experience), whereas the technical team, who were managing the data, technological and quantitative aspects, were primarily drawn from computer science-type backgrounds.

For the second organisation, the media company, the technical team were primarily working with machine learning algorithms and big data sources, and as such had deep skills in both the technologies and the quantitative approaches this entails. Their business-facing team were predominantly former consultants, mostly who have some quantitative background (the interviewee in this role had an MSc in OR). The benefit sought here was that although these people were not involved in much of the analytical processes, they had sufficient understanding to be able to translate customer requirements into technical problem spaces, and to provide an 'overlap' between the two functions. Indeed, this approach may be particularly relevant to machine learning and data mining orientated functions, where such approaches are more likely to entail greater programming and IT skills.

Again, this method is particularly relevant as it combines the benefits of the 'unicorn' approach but without necessitating the full range of skills this entails. Ultimately the individual teams in this approach only are required to have deep skills in two of these areas (technological and quantitative; or quantitative and decision making). The major drawback is that such an approach may be more relevant to the more data-rich organisations and potentially dependent on the successful collaboration of the two separate teams to be able to effectively implement and, moreover, operate. If there are communications blocks between the business-orientated and technical then the models may not adequately resemble the requirements of the client or the specific situation. For OR as a discipline, there may be an additional concern. If its graduates are better-aligned with business-orientated roles, as opposed to the technical roles, the quantitative aspect of the discipline may become less relevant and utilised.

6.3.10 Conclusions

The template analysis, as presented in this chapter, adds significant depth to previous analyses. As can be seen in the progression from initial to final template, as well as some of the discussion in the chapter, many of the concepts are consistent with *a priori* theory (i.e. the interview topics), and build upon earlier insights in this research. However, there were several emergent themes, particularly the five 'approaches' to structuring analytics teams, and the effect of this on skill requirements. To explore this further, and to evaluate some of the individual cases in a little more detail, the next section presents the results of the matrix analysis.

6.4 Matrix Analysis

As discussed in section 2.6.3, matrix analyses are, in many ways, a natural conjugate to a template analysis. The combination of the two, has an additional benefit of providing a more in-depth investigation of individual cases which can complement the more generalised results of the template analysis (particularly in the form it has been used here).

The central instrument is a $n \times m$ matrix, where n is the codes/topics under investigation, and m is the cases being analysed. For the choice of m, whilst in most matrix analyses the full set of cases would be used, for this analysis we decided to focus on one case to represent each of the five 'approaches' described in section 6.1.9. This is a more reductionist approach than is generally advocated in the literature. However, it is not inconsistent to the 'centre-right' philosophy of the research (as detailed in section 2.1) whereby some degree of generalisation is sought from these results, and such a reduction allows for a more manageable set of data for the matric whilst still maintaining a key structure in the data (the five 'approaches'). This analysis is supplemented with additional information on the additional cases in appendix item E. To focus upon the key requirements for each approach, the items included for n are the template items one to five (the 'skills' topics), and item seven ('outputs').

Doing so not only provides some case-by-case detail, but also allows a better understanding of the impact of these different approaches on skills requirements. A brief comparison of each of the eight template topics follows this section, whilst the full matrix is shown as table 32.

6.4.1 Data and Data Management

In comparing the use of databases across the cases, quite a significant range can be seen in respect to skills requirements. At one extreme, a consultancy described as using the 'on demand' approach, there were no requirements to this regard (as was also the case for several others who broadly employed this approach).

At the other extreme, the media company classed as employing the 'technical + business' approach, a great deal of emphasis was placed on such skills, with the team managing its own big data lake as well as using various relational databases. For the only other company classed as using the same approach, whilst they were not using big data sources and databases, there was a definite requirement for strong relational database skills.

The cases representing the 'devolved' and 'unicorn' approaches also had a relatively strong emphasis on such skills. One interpretation of this, is that considering that both require a stronger element of self-reliance, that there was less collaboration with other departments or resources, such that these abilities would be more important. For the final case, the travel company employing the 'operationalisation' approach, the need for such skills was not as greatly emphasised, but SQL skills and a basic awareness were deemed a necessity.

In respect to data used, all of the cases made some requirement for an ability to work with data of different types, and to be able to clean and find structure in them. However, for the examples of the 'unicorn' and 'technical + business' approaches this was emphasised more so than the others. It is worthwhile noting, however, that evaluating across the interviews this seems more a factor of industry and company type than of approach used. For example, a consultant considered to be using the 'unicorn' approach made it a requirement in client contracts that they took responsibility for provision of data in a broadly structured state, even though the company required staff to have significant computing skills.

Table 32 Matrix of skills requirements

Approach	Respondent	Data & Data Management	Quantitative Methods	Programming & Software
The 'devolved' approach	Digital Analytics Consultant	An understanding of databases, data warehouses and OLAP cubes. Abilities in business intelligence and reporting.	Abilities in data mining on large datasets. Ability to identify anomalies (errors/outliers) in datasets. Ability to add context to numbers.	• Business objects; • Web analytics tools (e.g. WebTrends); • SQL; • JavaScript; • Excel. A general awareness of programming rather than in-depth - enough to speak to specialists.
The 'unicorn' approach	Government Data Scientist	Working with structured *and* unstructured data), as well as multiple sources. Ability to work with databases. Able to sample from data.	Abilities in data mining and text mining. Abilities in statistics. Preferably some OR and/or maths.	• Statistical software (e.g. SAS); • SQL; • HTML5; • JavaScript; • Excel. "[A] familiarity with programming [and an] ability to play around with large data sets"
The 'on demand' approach	OR & Analytics Consultant	Database skills often required in projects, but are outsourced. "I absolutely understand that data is very good. But I've no aspiration that we are seen as data people really [...] We hope we've got a common enough language with people who *are* responsible for data"	Familiarity with OR (optimisation and simulation) and general model building skills.	• VBA; • Java. "We find that quite hard because we can't just ask people on day one, joining us, to [program in Java]. So we actually use associates [outsource] sometimes for that".
The 'operational-isation' approach	Analytics Manager (Travel)	Requirements for database skills and awareness. An ability to evaluate and clean data. An ability to evaluate the structure of different data types (such as web data).	Abilities in problem solving and modelling. An ability to design experiments and statistical tests. An ability to add context to data.	• Statistical software (e.g. SAS); • Business objects; • SQL; • VBA; • Excel. Moreover a "knowledge of the programming methods" than in-depth coding skills.
The 'technical + business' approach	Media Company Analyst	Requirements for Big Data and relational database skills (the team have built and operate their own data lake).	Requirements for machine learning and statistics, and to be able to write bespoke data science algorithms.	• R; • SQL. "[The team are] doing all machine learning type stuff, coding, but they've got the computational side of it [and] they've got a mathematical side of it."

Table 32 (continued)

Approach	Respondent	Soft Skills	Other Skills	Outputs
The 'devolved' approach	Digital Analytics Consultant	A basic requirement for some communication skills ("everybody needs those!"). More emphasis on visual communication.	No formal project management employed. An ability to convert analytics outputs to strategy recommendations.	Ability to create and interpret dashboards. Visualisation an important skill, though "often the best visualisations are the very simple ones". Ability to create effective presentations.
The 'unicorn' approach	Government Data Scientist	An ability to understand business context. Visual communication key.	Some project management (Prince2 desirable). "How to deal with stakeholders and what does a project look like. How do deal with milestone risks".	Abilities in data visualisation, "a lot of the stuff we do tends to be data visualisation". Some requirement for workshops and presentations, and reports/inserts for publications.
The 'on demand' approach	OR & Analytics Consultant	Soft skills seen as very important, and can be lacking: "in the case of two [recruits] to be really useful to us they needed time to develop confidence and, in a way, these soft skills"	No formal project management, but an ability to work in a 'agile' way.	Working model outputs, Excel-based data outputs and presentations and workshops.
The 'operational-isation' approach	Analytics Manager (Travel)	Problem solving skills the main need: "the people who can't solve and structure a problem are the same people who can't communicate it". Soft systems methodology is desirable.	Agile methodology used, and change management important, but experience not a pre-requisite.	Report writing and presentations . Visual skills required in terms of charts and graphs. "90%" of analyses incorporated into enterprise tools.
The 'technical + business' approach	Media Company Analyst	Technical team is not client facing, so limited skills required. Separate team manages scoping and dissemination. Some visual skills needed.	Agile methodology employed.	Presentations and workshops given by the client facing team, sometimes assisted by analysts. Data visualisation a key output and skill.

6.4.2 Quantitative Methods

Differences in the use of quantitative methods were less extreme than for the data management topic. All the cases required a general quantitative background, and an understanding of statistics.

One observable difference, supported in other interview cases from each approach, was for a greater emphasis on OR and modelling skills in the cases exemplifying the 'on demand' and 'operationalisation' approaches. In contrast, a need for data mining was more often cited for the other three approaches ('devolved', 'unicorn' and 'technical + business') in both these examples and elsewhere in the data. For the company representing 'technical + business', machine learning was the most important method.

There is some theoretical sense to this, in relation to the previous section. With the cases aligned to the 'on demand' and 'operationalisation' approaches demonstrating less of an emphasis on databases and data management, a lesser emphasis on data mining would be expected. After all, data mining is often described as one of the steps in the knowledge discovery in databases (e.g. Fayyad *et al*, 1996) process, (although, in practice, the terms are often used as synonyms).

6.4.3 Programming and Software

The use of programming languages and software packages, again, shows less obvious variation than for data and data management. Obviously, the cases who had requirements for database skills (all except for the 'on demand' example) also had requirements for SQL skills. VBA featured for both the cases selected to represent 'on demand' and 'operationalisation' approaches, but not for any others (although many still emphasised a use of Excel). The case that had the greatest emphasis on programming skills was the media company ('technical + business'), emphasising both their use for quantitative analyses (primarily using R), and for more general computing, as well as a focus on developing bespoke algorithms.

6.4.4 Soft Skills

As with data and data management, this topic saw relatively high variation, again with the cases representing 'on demand' and 'technical + business' at either extreme. For the former, soft skills were very important, albeit something recent graduates lacked (in meetings with clients at least).

For the media company ('technical + business'), the main analyst team had no major requirements for soft skills in new recruits, as a separate business-facing team managed these interactions. Some of the team act as support in certain presentations, but no pre-requisite is made. This was similar for the other organisation classed as 'technical + business' (not included in this analysis), and to some extent for the digital analytics consultancy ('devolved' approach). For all three organisations, a greater emphasis on visual communication is made, but typically in terms of visualisation systems and/or or dashboards (and typically delivered online).

For the government team ('unicorn' approach) and the travel company ('operationalisation' approach), again there was something of a mix. Both saw soft skills as important for delivering results in many situations, though other scenarios would not require client interactions. The former made emphasis on visual communication, but ideally sought recruits who were self-sufficient, although a "very important thing is [not to] be afraid to ask people [questions]". For the interviewee representing the 'operationalisation' approach, many (but not all) of their projects required scoping with clients, and advocated using the soft systems methodology. For them, the key soft skill was an aptitude to 'solve problems', from which they felt all other aspects were dependent on. Also, for reasons discussed in 6.2.6, an ability to communicate with the company's IT team was also key.

6.4.5 Other Skills

The main recurrent element from the 'other skills' category was project management. For two of the respondents, the companies representing the 'devolved' and 'on demand' approaches, no formal methodology was in place. The government department ('unicorn') and travel company ('operationalisation') both used a specific methodology (Prince2 and agile respectively), but neither saw this as a significant pre-requisite, and moreover a general awareness was a 'nice-to-have'. Only the media company ('technical + business'), regard this as a key skill (though it is worth noting the other company regarded as using the 'unicorn' approach, an analytics software vendor, was perhaps the most effusive about the importance of project management).

6.4.6 Outputs

Outputs, again was an area of some variation. For the companies representing the 'devolved' and 'technical + business' approaches, presentations were discussed, but typically recruits into technical roles were rarely involved in their delivery. Instead, both emphasised computer-based visualisations and dashboards. For the other three, presentations were a common output, alongside reports and workshops.

The government team ('unicorn') also spoke of the need for analysts to contribute to official publications, often in the form of tables, graphs and written segments. The OR & analytics consultancy ('on demand') discussed the merits and pitfalls of embedding working models into client teams, as well as a relatively extensive use of Excel-based data outputs. However, it was for the travel company for whom this section was the point of greatest difference, with 90% of their work operationalised into enterprise systems. This meant not only an importance on being able to effectively communicate with the relevant IT teams, but also the ability to produce pseudo-code representations of their analyses the teams could adapt.

6.4.7 Conclusions

Overall the matrix analysis adds to the earlier template analysis by ‘joining up’ some of the different topics around five of the cases to provide a clear picture of their skill requirements and organisational practices. Moreover, the analysis also offers some suggestion of how the different internal ‘approaches’ may impact these. Crudely speaking we may see these as a range where one end champions modelling, problem structuring and communication skills, with VBA as the main technology used; and at the other data management, data mining, data visualisation, and a wider range of programming languages and tools. This is represented in figure 40.

Figure 40 Skill requirement differences by internal approach employed

The two companies listed at the ‘people based’ end of figure 40 both had a clearer OR connection, whereas the two on the other end were more aligned to computer science, data mining, and, in the case of the ‘technical + business’ case, machine learning. Unsurprisingly, the case defined as using the ‘unicorn’ approach is situated in the middle, as recruits tasked with working on most of the tasks identified with analytics will obviously need a more even balance between both people and technology skills. Broadly speaking, the above statements approximately hold true for the other cases in the full sample, aligning to some extent with one or two of these approaches. Some evidence of this is given in appendix item E.

6.5 Summary

In summary, this chapter has provided in-depth analysis, complementing chapter four, of the skills requirements of analytics employers. To achieve this both template and matrix analyses were applied to the first group of interview data, and some brief discussion given.

Inevitably, there are some limitations. Without any clear or easy way to determine the proportions of the analytics community which are aligned to OR in comparison to machine learning or any other related discipline, it is hard to determine what a ‘balanced sample’ may look like. However, it seems reasonable to assume that there may be some over-representation of OR-affiliated respondents, as discussed in the methodology (section 2.6). Additionally, although there is perceived value in identifying the archetypical approaches presented in the previous section (6.4), the extent to which these results can be generalised is open to debate. Philosophical concerns aside, methodologically the approach is obviously reductionist, and the five approaches do not capture all the nuances in this sample alone. In combination with possible issues regarding sample size, this aspect in particular should be considered moreover as a working model than as a confirmed and comprehensive description of how analytics is structured across *all* organisations.

To help counter some of these issues, these results can be synthesised with those of chapter four and the earlier literature. This synthesis is presented in the discussion and recommendations chapter (eight), and thusly address research objective three ("To determine the skills requirements of analytics roles and the extent to which these may be met by OR professionals"). Prior to this, however, and utilising the final template presented here, chapter seven will apply similar analysis to the interviews conducted with academics and university course designers.

7 PLOTTING THE COURSE: CRITERIA, CONCERNS & CONSTRAINTS IN ANALYTICS CURRICULA DEVELOPMENT

The research thus far has covered definitions of analytics and its development; the historical development of the field and its relationship to OR and other disciplines; the key skills in analytics job adverts; made comparison with analytics degree materials; and, in the last chapter, captured some of the requirements of analytics employers. In doing so, we have accumulated much of the evidence needed to reach the principal goal of the research, to determine how analytics and OR degree curricula may be shaped to meet the needs of industry. However, to do so, one last element is required, a furthering of our understanding (building on the work of chapter five) of what is currently provided, and also some of the barriers and issues that impact curricula development.

In order to do so, as detailed in the methodology, we employed further interviews with academics and course developers working involved in analytics degrees as the last of this research's empirical studies. This chapter presents the results of this work, by analysing them using the template of chapter six. However, as not all the elements relevant to employers are necessarily relevant to educators, the analysis focuses on certain elements. These elements, which are discussed in sequence are: data and data management; quantitative methods; programming and software; soft skills; education; and, finally, trends in analytics. Prior to this, the chapter begins by evaluating general attitudes to analytics from with the sample.

7.1 Interview Methodology

As detailed in section 2.7, the academic interviews are designed to meet the following goals:

1. To contrast with the results of chapter six (employer interviews).
2. To be the compliment to the analysis of chapter five (the online course material analysis), towards understanding how analytics degrees relate to those of other disciplines (RO4);
3. To identify potential barriers for the development of analytics curricula (RO6);

The population under investigation is effectively academics working on analytics degrees from across all UK and Irish universities. From this population, we reached a sample of 13 participants in 11 interviews (one interview included three participants), and 15 questions were asked across a range of topics including definitions of analytics, core skill requirements, pedagogy and potential barriers for curricula development. A summary of the interviewees and the names with which they are referred to in this chapter, separated by institution location, is show in table 33. As in chapter six, our principal analytical method is template analysis, building on a version of that used for employer interviews.

Table 33 List of participants by location of their institution

Location of Instituition	Participants
Ireland/Northern Ireland	Professor; Associate Professor
Midlands	Associate Professor
Midlands	Lecturer
North East	Co-Director; Professor; Lecturer
North East	Emiritus Professor
North West	Head of Department; Emeritus Professor
Scotland	Professor; Senior Lecturer
South East	Reader

7.2 Attitudes to Analytics

As detailed in the introduction, this chapter begins by polling some of the attitudes to analytics demonstrated in the sample. In general, unsurprisingly, all participants were positive about analytics, and, in particular, the attention it was garnering in the academic community and in the more popular press. However, there were some different perspectives offered on the degree to which analytics represents something "new", or whether it is just business-as-usual.

> **Lecturer, Midlands:** "I don't see analytics as being different [...] for me, it's a combination [...] optimisation is a part of it. Statistics, I believe, should form a part of that [...] business analytics is a new name, a new brand, because technology has definitely [moved on]."

> **Senior Lecturer, Scotland:** "Everyone in their areas are wondering what is analytics? I mean, it's not like there is a very clear-cut definition [...] It's not just in this country, but everywhere, the US, Europe, Australia, etc. In some ways, I see it as when big data came out in a big way in the last few years, it's become more and more of an important area. In a way, OR is seen as a good opportunity, because obviously analytics is not any different to OR. You use a lot of analytical techniques in the context of massive data."

> **Emeritus Professor, North West:** "We recognised that this word 'analytics' was gaining some currency. We didn't see it as a new thing in terms of what we taught, but we thought it may be a useful word to better describe and emphasise some of our programmes."

An alternative perspective, given in multiple interviews and reminiscent of the arguments made in chapter three, was of analytics as more of an evolution and/or composition of disciplines:

> **Professor, Ireland/Northern Ireland:** "It's really going up in levels. [OR is] LP and queuing and that kind of thing for operations. And management science it broadens in and takes account of behaviour, so your more like marketing and strategy and things like that. But the analytics role is more of the bigger projects in terms of where it's all being used. Kind of the very highest level of companies who are not as much concerned with operations [...] its strategic decisions and brings more to the multi-criteria area [...] visual systems and decision systems."

> **Associate Professor, Midlands:** "[Analytics] is more interdisciplinearan [than OR] [...] You do need to have an idea of some computer science/data science type issues in my opinion, particularly when it goes towards big data, which is quite focused around IT solutions with the speed of data, the volume of data and so on. And obviously you also need to know about the statistical kind of things. Or management information systems, again something which isn't considered in [OR]. So, less optimisation [than in OR], more statistics, more computer science, more of these other things."

However, some identified potential barriers in this regard, particularly when trying to align with expertise elsewhere in their institutions:

> **Associate Professor, Midlands:** "We can't easily use resources from other departments in the university because there is quite a high fee discrepancy [...] We were interested in collaborating with [computer science] but in the end, it wouldn't work [because of fee differences]. Also, we don't necessarily want to get too close to the computer science people let's say, because essentially that's our competition."

Reader, South East: "We're looking at doing joint programmes [with computer science], it just an idea right now [...] sometimes there are [barriers] but it can be done."

Finally, many pointed to a significant focus on application over theory, seemingly beyond that of OR and other disciplines, and particularly on business applications. Many incorporated other business modules as part of their degrees, even those that are not directly linked with technologies of quantitative methods such as marketing and supply chain. This position was summarised in one interview as:

Emeritus Professor (North East): "[The core skills have] got to be how you turn data into money, into value. And if you lose sight of that ... how can you help an organisation find value in their data?"

7.3 Data and Data Management

The topic of data and data management was seen as important by all, but the extent to which degrees and modules focused on these issues. For one university, the extent to which big data is entirely new was of some question:

Co-Director, North East: "A lot of things are not new. Large elements are not new. Maybe textual analysis of large data files is new but in my commercial life I had one hundred million records to deal with on a daily basis in my department."

Another included two specific modules on big data, however, "more concentrated on the theoretical aspect" (Lecturer, Midlands). None of the universities in the sample made extensive use of technologies such as Hadoop or similar big data architectures. For some, this was considered "out of scope" for the type of degree or module they were delivering. For instance, one respondent considered there to be a clear distinction between types of analytics, (echoing the findings of chapters four and five), and that big data's role was not necessarily the same within each, whilst another considered much of big data to be the domain of the computer scientists:

Lecturer, Midlands: ""[We] show interesting case studies [...] some big data examples. Not to show them a particular software [but] to demonstrate that nowadays, this is the case, that it's not just having one software and some data, it's a matter of having a big amount of data, it's a matter of having parallel platforms running together [...] All of these platforms, all of these things, they will be looked after by computer scientists. What we are aiming [for] is business people. They will become consultants, some of them will get senior positions. It's about the front level, how you interpret things."

Professor, Ireland/Northern Ireland: "Decision analytics is different to data analytics and to big data. So, its big data vs. small decisions [...] smaller decision because they don't require that much data, but they are likely more important [to the business]."

However, one respondent suggested this was also a skills issue:

Associate Professor, Midlands: "[Big data] is something we don't do in much detail. We do talk about big data, explain what it is [...] but it's mostly about the business understanding at the moment, about how can we innovate [...] Big data also comes down to expertise, we wouldn't have anyone who is an expert in Hadoop."

An area of significant concern in several of the 'employer' interviews, was the use of artificial datasets and exercises, deemed unrealistic of real world analytics where problems are typically ill-defined, and where datasets are very messy and noisy. Interestingly, this issue was recognised by all interviewees and all reported *some* use of realistic datasets, with one arguing this is a key differentiator from an OR course for instance:

Associate Professor, Midlands: "Working with real data, that's really important. [...] Profane tasks like data cleaning [...] used to be not taught in our degrees because we'd be teaching methods and techniques [...] Analytics is more about how to deal with data."

It was noted, however, that 'realistic' is not necessarily the same as 'real':

Senior Lecturer, Scotland: "We do have a lot of practical classes with genuine clients coming in, giving lots of messy problems. But, I would say artificial data comes in handy most of the time for the data itself. It's often quite tricky to get numbers from a client but they can come in and maybe talk to students and give a bit more qualitative information. That's why artificial data creation comes in handy. For example, if you wanted to give them big data or messy data then you usually create it yourself. [For a module] I created some data which was messy enough, realistic enough."

For others, data cleaning and 'messy' datasets were appropriate in some parts of the course, but not necessarily used in every instance:

Associate Professor, Ireland/Northern Ireland: "I think that it depends on what I'm teaching. I mean if I'm doing a data mining course I probably want to be dealing with dirty data because it is such a big topic. But if I'm teaching theory of regression or something, that's absolutely no reason we should spend the first half of the class teeing up the data."

Lecturer, Midlands: "The module [I am running] is more theoretical. I think artificial data is suitable for this. For something [...] related to applications, [real data] will be of benefit."

In the same spirit, some saw this as more of a progression, where students are first introduced to techniques in a more controlled fashion, and therefore using artificial data designed to be used for the task in hand. This can then be followed by modules or projects where students are tasked to try to apply the approaches learned in more realistic settings where there can be multiple ways to address the problem, and where datasets are more realistic and messy:

> **Emeritus Professor, North East:** "We have [two modules]. Session one is the basics and session two is applied. So, there we'll get real data from local businesses, a local company or a local organisation, and have the students apply the skills they learn."
>
> **Head of Department, North West:** "Almost all the Master's projects are with a company [...] These are real-world problems and data with an industry partner [...] In the modules its [artificial] datasets we have, because you can't do it. Real datasets are messy [...] it doesn't really work in a classroom [setting], you'd spend half the time understanding the data."

Consequently, there does seem to be some disagreement between what employers are seeing in graduates, and the opportunities (purportedly) made available in the courses, with all using *some* real or noisy datasets at some point. There are several possible explanations. One may just be that despite some training, this is an area students can still struggle with. This may suggest that current training is insufficient and would either need to be of greater quantity or greater clarity. Alternatively, this may be an effect of the sampling used; that the universities consulted in this stage of the research are not representative of all UK universities, or that the 'employers' consulted have either been unfortunate in their hiring, or have hired from other courses which are not offering such opportunities to work with real data.

A similar contrast can be seen between the emphasis many 'employers' made on data management and associated skills and/or software, and the inferred coverage of the topic in the courses of the institutions in this sample. For instance, one employer considered the acquisition of team members with database skills, to bring about something of a "revolution" in their department (section 6.3.1), whilst several others highlighted many challenges they were facing in dealing with big data projects and unstructured data. This emphasis is not seemingly matched by the provisions of the degree courses investigated in this part of the research. Almost all acknowledge some importance, and cover some aspects in their curricula (e.g. data cleaning or some theoretical aspects of big data), but there does not seem to be direct parity.

7.4 Quantitative Methods

As would be expected, quantitative methods were a major element of all the relevant courses provided by participant's institutions. A summary of the areas included (utilising the template 'child nodes' shown in figure 31, section 6.3.2) in each of the institutions in the sample is shown in table 34. It is worth noting that this list relates more to the importance of the methods in their courses suggested the respective participants, than it is a definitive assessment of their university's coverage, as many courses had *optional* modules in other areas, as well as the list being those they chose to highlight in their responses. However, it does provide some indication of offerings.

Table 34 Coverage of quantitative disciplines across the institutions in the sample

Location of Instituition	Participants	OR	Data Mining	Machine Learning	Statistics
Ireland/Northern Ireland	Professor; Associate Professor	✓	✓		✓
Midlands	Associate Professor	✓	✓		✓
Midlands	Lecturer	✓	✓		✓
North East	Co-Director; Professor; Lecturer	✓	✓	✓	✓
North East	Emiritus Professor	✓	✓	✓	✓
North West	Head of Department; Emeritus Professor	✓			✓
Scotland	Professor; Senior Lecturer	✓	✓		✓
South East	Reader	✓	✓		✓

In all interviews OR (particularly optimisation), statistics, and data mining were discussed, with the former two seen by many as core components:

> **Professor, Scotland:** "[Analytics courses are] expanding the boundaries of OR to take in more stats and maths."

> **Associate Professor, Midlands:** "We won't start bringing in optimisation techniques till fairly late. We'll start with statistical techniques on how to get insight from the data."

The frequency with which data mining is included is of some interest. Indeed, for one participant, this has the consequence of limiting traditional OR techniques:

> **Professor, Ireland/Northern Ireland:** "We would have had a lot more traditional OR subjects [when we launched our analytics degree]. Now we have Java and different kinds of programming, and we still have statistics. Not as much simulation and more data mining."

Other subject areas mentioned included forecasting, the next most frequently suggested, and other approaches such as revenue management and text analytics were mentioned in individual interviews. One element that may have been expected, based upon the literature review and 'employer' interviews, was machine learning. However, this was a key component for only one of the interviewees. Indeed, another argued:

> **Head of Department, North West:** "[Machine learning] is more computer science [...] That's more 'how do you understand the data, how do you structure the data, what can you learn from the data'? I think that's more computer science."

The absence of machine learning components may correlate with the relative absence of practical big data coverage, as the two are often used in conjunction. Big data, however, was seen by more than one respondent to present both opportunities, but also invalidated many of the traditional analytical approaches. Such statements problematise the comparative lack of coverage on analytics on big data sources in the institutions included in the sample:

> **Head of Department, North West:** "It has to be different [algorithms for big data] because of the amount of data. Lots of algorithms have been developed in [OR] over many years [...] Now it's different, now its driven by the data [...] Many traditional techniques cannot cope with the amount of data."

Despite such questions, however, there was clear consensus of the importance of quantitative elements. Indeed, for one respondent this was perhaps the critical element in an analytics degree, with other aspects easier to pick up later:

> **Emeritus Professor, North East:** "It's much easier to train an [analytics professional] programming than it is an IT person [to learn the quantitative elements]. The maths and stats are quite tricky. I think anyone who can do good stats, will learn to program [...] It will be harder for a programmer [to learn analytics], because you'll need a 5-day course just to get the absolute basics and you'd be talking about at least 20 days' training to become any good - it just takes a long time."

7.5 Programming and Software

The previous topics, *quantitative methods*, showed a reasonable level of consistency overall between respondents. Programming, however, was an area of some discrepancy between institutions. It is therefore worthwhile, before addressing these, to revisit the findings of the template analysis of 'employers' perspectives on programming requirements.

The most 'in-demand' of these was found to be SQL, followed by Java, VBA, SAS, Python and R. Other software such as SPSS, AIMMS and Simul8 were also mentioned, but to a lesser extent. Perspectives on the importance of programming and software also varied. At one extreme they were considered critical, and, in the words of an Analytics Consultant, skills that without which "no-one would stand a chance out there" (section 6.3.3). At the other, it is seen as something important to working in analytics, but not something a graduate need come in with; moreover, something that can be taught in the job (section 6.3.3). Against this backdrop it is perhaps not surprising that universities too weight the importance of programming with different degrees. In some programmes, it was considered beyond the scope of what they were trying to achieve:

> **Associate Professor, Midlands:** "We're using, at the moment, SAS Enterprise Miner and SPSS Modeler as data mining tools [...] They are visual tools. Since we've only got the one year available it seems not really realistic to teach them something like R for instance, which would probably be quite useful, but takes a while to learn the syntax [...] What's more important is that by using the software they are getting to learn the overall process."

At other institutions, the value of programming was held in much higher regard:

> **Emeritus Professor, North East:** "[Analytics is about] the maths and statistics, [but also] definitely programming [...] It's no good being good [at the] theoretical, if you can't get your hands on the data, manipulate it, and do something with it. So really its learning to program - SPSS doesn't cut it. So, you need R or Python, scripting languages."

Scripting languages was also a point of debate for another institution, one which also seemingly considered programming a key skill:

> **Associate Professor, Ireland/Northern Ireland:** "We teach Java in one of our courses and actually in two of our courses, one of them is optional. And, you know, that would not be my first choice as a teaching language. I think some of the students kind of struggle. It's a bit heavy for what we want to do. You could be teaching them Python, which is very lightweight by comparison [...] If you look at a certain type of job ad it's more likely to say Java than Python. And that was certainly true five years ago. [...] Java is a compiled language and once we've exposed students to that then it's probably easier for them to move to a scripting language. [...] Once you've got the idea of what compilers do then you can probably figure it out, how to work without them. But, if you go in the opposite direction then there's a bit more of a learning curve."

Between these perspectives, there was of course plenty of middle-ground. Some institutions offered single solutions (SAS being the most popular seemingly) and others taught multiple languages/software in their courses across different modules:

> **Head of Department, North West:** "There are all sorts of software on different modules, its module specific. So, when I teach optimisation I use Excel [...] some modules involve SAS [...] we've introduced a new module on enterprise systems which uses [...] SAP [...] we also have C programming."

> **Reader, South East:** "We teach some software, and I think we should teach a bit more. Right now, we have SPSS, MySQL, [...] Visual Basic Applications, Simul8 [...] But I'd like to do something more around big data [languages]. We aren't doing that now."

In terms of what motivates language or software selection, several influences were identified. As, arguably, the motivation one may expect to be the chief concern, several pointed to a desire to best equip students for the requirements of employers:

> **Professor, Ireland/Northern Ireland:** "We just try to use the tools that are being used in business."

Besides this, other factors are seemingly of import, such as the ease with which software can be distributed, partnerships with software providers, or the perspectives of other stakeholders:

> **Senior Lecturer, Scotland:** "I like to use FICO Express, mainly because I can easily install a dynamic license on the server. That's one of the major things for me, I basically install it on the University server and then there are five hundred licenses available to students and all our students can use it and I don't worry about whether they are installed properly."

> **Professor, Scotland:** "We have an advisory board which we meet with once a year and the MSc is high on the agenda to talk about every year. It's made up of heads from OR and every year we say to them, 'is there something that we're not teaching'? And there's never really anything significant that comes up, because often it's more software orientated [...] They still want us to teach stats but they might want us to upgrade to a different software."

However, even the two participants who had the highest emphasis on programming in their courses both stressed that students need not obtain complete mastery of any specific languages:

> **Associate Professor, Ireland/Northern Ireland:** "So you can learn to use data mining algorithms, you can learn RapidMiner [...] From the point of view of a user and I think that's what the students often think they want. But the other part of our philosophy is that that's not what's good for them. We'd better give them the theory and the foundations, the mathematical foundations whether they like it or not. And then they can go and pick up the practical tools to some extent on their own."

> **Emeritus Professor, North East:** "You need an ability to learn really quickly [...] I don't write code. I just search on all the blogs and [...] somebody's written it [...] That was true of COBOL when I started in the '70s. No-one wrote programs from scratch, you took someone else's and you modified it [...] You have to know enough to be able to modify the code, if you don't understand what you are doing - you might be lucky and it will work - but often [...] There's a [software library] for everything [...]I think the core skill is the bricoleur really, the ability to pull packages, work with different software, go out and find stuff. And you don't have to be a great programmer, you just have to be good enough. If it does end up in enterprise software, then you need IT professionals. That's not [your] job."

7.6 Soft Skills

As with the 'employer' interviews, there was significant discussion around soft skills and their relevance to analytics:

> **Associate Professor, Ireland/Northern Ireland:** "The philosophy [of the course] is that we go beyond the [the stereotype of] the nerd in the corner who is really good at algorithms but he doesn't know how to consult with clients or communicate with other co-workers or deal with management or you know translate it into business outcomes."

> **Associate Professor, Midlands:** "Understanding the modelling techniques is very important, but likewise its important, possibly even more important, to have a good understanding and feel of the business model, what are the business questions?"

In this regard, many highlighted the value of OR, particularly the UK interpretation of it:

> **Head of Department, North West:** "UK OR is very different from elsewhere [...] One of the most important contributions from [UK universities] is the soft systems."

> **Professor, Scotland:** "We need these more rounded skills to send them out there [...] It is about client engagement and OR softer skills that entail engaging, gaining results and structuring a problem and understanding what's going on."

The delivery methods, however, were varied. In some cases, specific modules were offered:

> **Lecturer, North East:** "We've also included Psychology [as a module] because one of the things you have to be good at, especially when you are dealing with managers who aren't particularly good at analytical approaches, is persuading and putting together a good case for understanding how they make a decision, where it goes wrong and how you can help them make an informed decision using all the information. Because, the best analytics in the world is no good if people's judgemental biases prevent it being useful."

> **Associate Professor, Ireland/Northern Ireland:** "[We have a module] about consulting with stakeholders and figuring out what people are trying to get from a decision, their priorities and it's much less quantitative [...] There's a second module that's optional in the second semester and again that's about decision support systems, which also talks about that type of material but in more of the context of decision support where you have some sort of quantitative tool but it feeds into a human decision."

In others, these elements were included alongside other parts of the curricula. Several included modules designed to simulate real projects and many sought to develop this through consultancy-type projects. In around half the institutions, a summer project of this type was included with real-life clients, something highly sought after in the 'employer' interviews. However, even when real clients are not available, some identify ways to add this realism:

> **Professor, Ireland/Northern Ireland:** "[On a module if we] don't have enough [real world] projects and are running short, then there is a standard career one. In a group of four, two people [in the group] who won't know what they want to do next year in terms of their career, what kind of job they will be getting. So, the other two actually have to build a multi-criteria model, as a consultant, to solve their problem [...] These people who seriously want to know whether they should move to the States, get a job, set up their own company or take a year off and go travelling [...] They really want to know the answers. So, the main thing is to represent a real-life situation, where people will be engaged."

> **Emeritus Professor, North East:** "We partner with an incubator [...] and one of the things I want to do is get the students down there for a day, and have one of those hackathons, where they get given a real dataset and a problem [...] So, what we're trying to do is come up with something practical, realistic in terms of something people really do."

More specifically, several of the individual skills discussed in 'employer' interviews were also identified as components of these courses. These included change management, project management, negotiation, sales skills, and consultancy. All interviewees considered it important for graduates to have an ability to communicate and collaborate with end-users, business decision makers. However, one participant also highlighted a need to communicate with other parts of the organisation:

> **Associate Professor, Midlands:** "We need to teach students to be able to function in a team, because analytics projects you should have on a team somebody who is from the IT side, and you should have someone from the business side, and you should have someone who can connect the two of them. Otherwise people will be clashing all the time. Nobody would know what the other is talking about. So, this is kind where we see [our graduates], not being specialised [in IT *or* in business], but [able] to understand the language of both."

7.7 Education

Unsurprisingly, education was a key topic in these interviews, even more so than in 'employer' interviews. Our sample is primarily drawn from what we have called 'type two' programs (section 5.3); courses typically based out of business schools and with stronger associations with OR than machine learning. This has some limitations, which will be discussed later in the chapter, but does not prevent us seeking to explore perspectives on 'type one', data science and machine learning orientated programs:

Professor, Ireland/Northern Ireland: "A lot of our competitor programmes probably are the Computer Science Masters degrees in Data Science and similar. I think that a lot of the stuff is very similar [...] if you talk to an employer they probably don't distinguish that much between the two [...] From looking at other programmes, like Data Science programmes that it would tend to be a bit more practically focused. So, they will look more at data, big data and practical tools and maybe less mathematics and less of the business."

Some of the participants stated a desire to include new content that may be more associated with such programmes, particularly around machine learning and analytics on big data:

Reader, South East: "There's such a variety of topics [in analytics], you can't offer all of these modules. And there is resource restraints, though the [school] is quite large [we would like to do] text mining, for example, social media analytics [...] but we'd need to hire more people with that kind of experience."

Multiple respondents found recruitment of staff to be non-trivial, and highlighted potential issues with recruiting academics:

Professor, Ireland/Northern Ireland: "We can get a lot of mathematicians and computer science people. But it's the bit that connects that and is about making the decision and connecting it to the business that's the difficult one. And part of it is that you have to have a PhD [to get the job]. Practice is what has generated analytics, not sciences and not humanities."

Associate Professor, Midlands: "As an academic you are trying to focus on a small field [...] that's the way to get published [...] Whereas in analytics [teaching] you need to be very broad across the spectrum."

Emeritus Professor, North East: "[Is there a disconnect between academic and practical analytics?]. Absolutely there is. Our journals value theory, not actual insights, predictions. All the top journals are obsessed with theory."

Many highlighted the importance of external, business partnerships to help counter this issue:

Lecturer, Midlands: "For a course to be very successful [we need to have] companies working alongside us [...] That will be much easier to run. It will be much more efficient. Less costly. And the students will have the chance [...] by discussing with industrial partners what's happening, to get to know before [...] about the [work] environment that follows. Because otherwise, you don't really know what's going on."

Another potential solution to such a problem would be inter-department collaborations (with a computer science department for instance), but, as previously discussed, this can present issues:

Senior Lecturer, Scotland: “When I was [at a US university], yes there were departments and so on, but then there was actually good motivation for you to offer interdisciplinary classes. You could easily run an OR course and people from Engineering would come join. I’m not saying this is not happening here, I’m just saying there nobody was worrying about we are offering this class under the Industrial Engineering Department when these other engineering students come in. So, we’ve got five students and five students is this much money and this department should get this amount of money, etc. I think those are the boundaries which actually hurt things. Then you are always thinking what do we get in this department versus thinking in the broader sense of what you are providing at the University as a whole.”

Alongside this issue, there were other potential barriers identified that can limit or restrict curricula reform:

Senior Lecturer, Scotland: “I think curriculum design is always messy. You decide you have 12 weeks, which modules you are going to do, which topics are most important. I see that in a class I taught. Not in a bad sense, but somebody designed the course and then they left. Suddenly I came in and was like ‘oh there’s this new class, this is the way this other guy designed it and I was not sure if I would have designed it this way’, but then that was already approved and I had to stick to that.”

Associate Professor, Ireland/Northern Ireland: “Any type of curriculum reform runs in to, well, ‘you can’t do that because the part time students won’t be able to go to that class’ or ‘they clash’ or this sort of thing. I see that as an obstacle now to any kind of changes to the course.”

Another issue highlighted, and one which hitherto had not been really considered, was awareness of analytics amongst the students themselves. From within the analytics community, and indeed much of the business community, there may seem to be a lot of ‘hype’, but to some extent this is an echo chamber, and potentially this message is not reaching potential students:

Reader, South East: “The difficulties in recruiting students is that ... Universities are aware of the need for producing students with analytics skills, for the students, analytics is still not a common enough, popular discipline.”

Associate Professor, Ireland/Northern Ireland: “There are students out there who would benefit from our Masters, and a lot of students who finish at undergraduate are somewhat unfocused about what they want to do next and if somehow our Masters came on their radar it might make a lot of sense to them because it would be a way for them to transfer what current skills they have into something very employable [...] There are

> probably some people, say Engineering, Physics, Chemist or other sciences like that and also people who have a bit of a Maths background or a pure Maths background who maybe don't really have an obvious vocation but they could use our Masters to transfer. So, part of it is just marketing. How do we ourselves in front of those people? It seems like everyone is talking about business analytics all the time if you listen. It seems like it's an echo chamber effect [...] Peers talk about it, and we read newspaper articles about it, but then you realise from the point of view of a student [...] they're not seeing that at all."

The final question of the interview asked respondents to consider the extent to which universities should seek to change based on the needs of business and the jobs market, over an importance in maintaining academic tradition. Some felt universities should be closely aligned to business need:

> **Reader, South East:** "Every university should adapt based on the needs of employers. The demand from employers. I think it will be a bit short-sighted to get rid of a program [based on student demand] ... it's unlikely that a specialist program will attract as many students as a generic program such as 'MSc Management', [but] they put so much emphasis now on employability, our students do [get jobs]."

However, an alternative view is that basing solely on current demand may not be in the best interests of students, and that a longer-term view may be appropriate:

> **Associate Professor, Ireland/Northern Ireland:** "We're definitely not just serving the needs of business. I think we are serving the needs of students but I suppose. In terms of our Masters I think I'm fairly happy in saying that our goal is to serve the students interest as we see them, not as students see them necessarily and not as the market sees them [...] After the student has been in the market place for ten years are they still relevant? Are they capable of independent thought which will make them stay relevant? Rather than what's going to get them a job for the next six months. [...] I'm not crazy about our academic traditions are sacred or anything like that but we should be making our own decisions as opposed as being slaves to the market."

A similar argument was made about a focus on immediate skills over underlying principles:

> **Professor, North East:** "I get nervous about students coming through the system without getting a serious foundation, in at least one of the traditional disciplines. Now, I' quite happy for those disciplines to morph over time and into each other and so on but that knowledge base and the ability to think and problem solve within that space is very important. So, I am really nervous about programmes that just superficially take you across a whole wide area without giving you the tools to think with."

> **Lecturer, North East:** "Is education about skills or underlying concepts and application of the skills? [...] I think in some ways, say someone is trained in SAS, then that's testable, that's quantifiable but the underlying concepts become harder, don't they? It's a bit of a double-edged sword in many ways."

Such debate highlights some of the difficulty in tailoring courses. On the one hand, there is some 'duty of care' to the students such that they get a level of education that can prepare them for the full length of their career, and to include elements which are not necessarily easy to acquire, or easy to objectively measure, such as underlying theories and principles. On the other, there is a need for courses to appeal to both students and to employers, both to ensure adequate class sizes and to encourage business participation (whether that be datasets, consultancy projects or strategic direction). In this regard, the more quantifiable elements (e.g. specific software certificates) and practical skills currently in demand would appeal.

7.8 Analytics Trends

Participants were asked, as 'employers' also were, about the trends they saw in the analytics space, both current and future. Some responses regarded specific techniques seen to be growing in prominence:

> **Lecturer, North East:** "I think that the next big area is probably going to be deep learning. So, that's just an adaptation of Neural Networks really [...] But, you can't use them for everything. So, an example, in Financial Services, however good the model you can't use it to make a credit granting decision because you can't justify the decision. You've got to be able to justify the decision."

Most others identified trends that were related to changes to analytics provisions in academia, many of which were the continuation of patterns that have already been discussed over the course of this chapter. However, most agreed that analytics would continue to grow within academia:

> **Reader, South East:** "A few years from now business analytics is going to be a discipline as common as other business disciplines."

If indeed such a rise were to occur, this may have relatively profound effects on some of the issues and barriers listed in this chapter. Most obviously, larger student populations would allow for the recruitment of more specialised lecturers (e.g. in big data architecture or machine learning), and would obviously mean that the awareness of analytics amongst students would be higher, and the programmes more visible.

7.9 Summary

Over the course of this chapter, several of the core elements of the template developed in chapter six (the ‘employer’ interviews) have been applied to the 11 interviews with academics working on analytics degrees (either teaching or administration). In doing so, insights, agreements and differences have been presented regarding how they define analytics; the inclusion (or otherwise) of the areas of *data and data management*, *quantitative methods*, *programming and software*, and *soft skills*; general aspects of analytics *education*; and *analytics trends*.

Over the course of this discussion, some of the similarities and differences with the results of the ‘employer’ interviews have been presented, and this topic will be evaluated further in chapter eight. Equally, the results will be synthesised with those of chapter five, the analysis of online materials associated with analytics degrees, to address RO4, an assessment of the relationship between analytics curricula and that of other disciplines (such as OR).

It is worth noting at this point, that the choice of sample may have some implications for this analysis. Primarily the focus has been on what we have labelled as ‘type two’ analytics courses (section 5.3); those that are typically based in business schools, have a clearer association with OR, and are likely to carry titles such as ‘MSc business analytics’. These were held in contrast to ‘type one’ courses, which are more likely to be hosted in computer science schools, have a heavier focus on machine learning, and to have titles such as ‘MSc Data Science’. In using this sample, a consequence of finite resource and access, there is an unwanted and unwitting limitation to the analysis which will need to be managed. However, as the overall research is more focused on the position of OR, and its influence on analytics, it is better this way round than the other, and the results still provide additional information towards meeting this objective.

Additionally, some consideration should be given to the degree of expertise participants have in analytics. To some extent, as had been alluded to earlier in this chapter and in the thesis as whole, an expectation for any one individual to have deep expertise in all facets of analytics may be unrealistic (or at least rare). Whilst this is likely true in any walk of life, in academia, as noted in one interview (section 7.7), there may be additional call for research specialisation. In other words, as far as research goes, the nature of how universities make academic promotions may favour breadth over depth. Potentially compounding this issue, with many UK analytics courses in their relative infancy, or in some cases more a ‘rebranding’ than new curricula, in some cases academics may be better described as “finding their feet” with analytics, than necessarily fully immersed in the field.

Finally, this part of the research was designed to evaluate some of the potential barriers to the development of analytics curricula (RO6). Whilst this has been presented across the course of this discussion, chapter eight will seek to summarise them and evaluate the impact they may have on future developments. Chapter eight will then conclude by synthesising the results of all the analyses discussed in this thesis, to present a suggested framework for both analytics and OR-type degree curricula, thus addressing the seventh and final objective of this research.

8 Extra Curricula: Implications for Analytics Course Design

Over the last five chapters, building upon the earlier literature review (chapter one) and methodology (two), a variety of research instruments are presented. Each of these instruments sought to meet specific goals and objectives, but in combination they also offer insight to our central problem space, the content of analytics and OR degrees in UK universities. The purpose of this chapter is to synthesise these results and insights, and in doing so to provide recommendations on how such curricula may be developed to best meet the needs of employers. Through this process, we not only meet the final research objective, creating a framework for the development of analytics courses, but also, in doing so, address the central aim of the research; an understanding of how OR and analytics courses can develop graduates equipped for a career in analytics.

To this end, the chapter is arranged as follows. Firstly, a summary of the results generated by each research instrument is presented. Thereafter, the analyses of chapters four and six are synthesised to help define the requirements of analytics employers (research objective three). Thirdly, the insights from chapters five and seven are combined to address objective five, and to determine the de facto 'as-is' process; the current provisions of analytics curricula. Finally, and in consideration of all of these elements, a set of recommendations are made for both analytics and OR curricula and teaching, and therefore addressing objective seven.

8.1 Summary of Previous Analyses

To begin the chapter, as indicated in the introduction, the results of chapters three to seven will be briefly summarised. Through reflecting on these, we provide the background insight that allow for the recommendations presented in the chapter.

8.1.1 Historical Analysis

Chapter three detailed the first of the research instruments, a historical analysis of, what was described as the *dianoetic management paradigm* (section 2.3). In this paradigm, logic and discursive reasoning, supported by technology, quantitative methods and decision sciences, has been the dominant approach to meeting organisational problems and to inform decision making. The analysis was designed to meet research objectives one and two: identifying the relationship between analytics, OR and other related disciplines; and formulating a research agenda for the OR community that can address the specific challenges of the analytics age.

In respect to the former, our analysis proposes that OR, and a range of other technological, quantitative and softer, behaviourally-orientated disciplines in fact co-exist in this paradigm. As such, OR, or any other related field, is both an independent line of enquiry to analytics, but at the same time will seek to inform, to borrow from, and to compete for customers (end-users in organisations charged with commissioning or executing analytics work) in the 'shared space' of analytics. In reaching this conclusion, not only do we reach a 'working understand' of how OR and analytics can be understood (at least in the frame of this research), but also suggest implications for how curricula should be designed, and therefore how our recommendations should be framed. In particular, an approach that this chapter adopts, it is necessary to make recommendations for both analytics degree courses, and also OR courses (as we recognise OR as being a distinct entity to analytics as well as a core component of its teaching).

Secondly the chapter presents a research agenda for the OR community that can meet the specifics of analytics' current concerns (section 3.4). While this is of less direct concern to the discussion of this chapter, nevertheless it remains an important consideration for research-orientated academics in the analytics field.

8.1.2 A Topic Model of Analytics Job Adverts

The second instrument, presented in chapter three, was an Online (batch-based) version of latent Dirichlet allocation, used to analysis the key topics (subject matters) present in analytics job adverts, and those of six related disciplines (OR, statistics, machine learning, computer science, information systems and psychology). The adverts were sourced from the popular online jobs board on LinkedIn.

The function of this analysis was to provide some insight towards research objective three (extended in chapter six, summarised below), an evaluation of the key skill requirements for analytics roles, and the overlap (or otherwise) with the skills that may be associated with OR professionals and graduates. Section 4.2 detailed a correlation analysis of the topic proportions generated. The analysis found a relatively high correlation between analytics job adverts and those of the other disciplines (as may have been expected). Correlation between OR adverts and analytics adverts was at 0.85 (third highest behind statistics and machine learning). Additionally, correlations were computed for topics associated with hard skills, soft skills and domains. Correlation was highest in respect to hard and soft skills (0.79 and 0.88 respectively, both second highest), however was weaker in respect to domains (0.75, 4th highest).

Based upon these relationships, and further qualitative analysis of the results, several comparisons were made between analytics and OR, and the skills most frequently occurring in the adverts of each. These were summarised in table 24, repeated as table 35.

Table 35 Comparison of skills requirements in analytics and OR job adverts

	Hard Skills	Soft Skills	Domains	Programming
Similarities	Analysis (quantitiative); software development; Big Data	Management (skills); communication skills; consulting	Marketing; financial (control); financial (audit);	SQL; C; C++
OR+	Modelling; machine learning; process monitoring	Project management	Manufacturing & SCM; intelligence & operations; engineering & safety	R; SPSS; Matlab
Analytics+	Programming; solutions & architecture; business intelligence	Analysis (business); sales skills	Marketing campaigns; ecommerce; advertising	Java; JavaScript; HTML

8.1.3 A Quantitative Analysis of Master's Degree Content

The third instrument, presented in chapter five, was again a quantitative analysis of text data, in this instance using online materials associated with degrees associated with analytics, and the six related disciplines from the chapter before. The analysis used a bagged support vector classifier (SVC), trained on the same job advert data from chapter four, to predict analytics degrees to one of the other six disciplines. Additionally, a new metric was developed, module topic weighting (MTW), to score the frequency of different modules in the different classifications of degree type. In essence, the idea was to find the disciplines with which these degrees were most associated (those they were classified as), thus meeting research objective four. The work also made contribution towards (along with chapter seven) meeting objective five, identifying the skills currently incorporated in analytics curricula.

The SVC favoured two classifications, OR degrees and machine learning, with 46.5% and 39.5% of the classifications respectively. Using the MTW scoring to further analyse these two categories showed those classed as OR were most likely to feature modules on finance, marketing, statistics, data analysis, and data mining. Those classed as machine learning featured big data, data mining, machine learning, web & eBusiness, and programming.

Additional analysis of the degrees found further patterns. OR classed degrees were more likely to be titled "Business Analytics" or similar, with "Data Science" or "Big Data Analytics" the most likely title for machine learning classed degrees. The former also was most frequently offered in business schools, and the latter from computer science and technology schools. Using these results, a typology was suggested, which infers a two-pronged approach to analytics curricula in the main. The two types identified is shown in table 36, a repeat of table 30 from section 5.3.

Table 36 The two types of analytics degree in UK universities

	Type One	Type Two
Likely title(s)	"Data Science" / "Big Data Analytics"	"Business Analytics"
SVC classification	Machine learning	Operational research
Likely school	Computing / Technology school	Business school
Linked disciplines	Machine learning (35.15%) Computer science (27.42%) Information systems (19.13%)	Operational research (38.72%) Statistics (24.90%) Information systems (20.17%)
Most likely module topics	Big data (9.46%) Data mining (5.99%) Machine learning (5.36%) Web & eBusiness (5.05%) Programming (4.73%)	Finance (7.20%) Marketing (5.46%) Statistics (5.21%) Data analysis (4.71%) Data mining (4.71%)

8.1.4 Interviews with Analytics Employers

Chapter six presented the second instrument employed to meet objective three, determining the skills requirements of potential employers of analytics and OR graduates. Complementing the quantitative analysis of chapter three, interviews with 29 potential employers, from a range of domains, were conducted, and analysed using template and matrix analysis techniques.

With a dataset of this size, there were several insights drawn from the data, around a range of relevant topics. However, arguably the most significant was that employers utilised a variety of approaches to structuring analytics teams, and that these differing approaches entailed different requirements. The five approaches identified were as follows:

- **The 'Devolved' Approach:** This approach resembled the idea of the virtual enterprise in that rather than there being fixed teams who managed analytics projects, the organisation (there was only one in the sample who fully employed this approach) was moreover a network of specialists who could be engaged as required. This effectively allowed for hybrid combinations of skills as the projects required.

- **The 'Unicorn' Approach:** In contrast, in this approach analytics specialists worked in a more end-to-end capacity, with responsibilities across all three of the main areas of the Venn diagram of figure 14 (section 2.3), in *technologies*, *quantitative methods* and supporting *decision making.*
- **The 'On Demand' Approach:** In this system analytics specialists were mostly required to manage the quantitative and 'front-end' activities (problem structuring, stakeholder management and similar). If skills using specific technologies are required, resources were employed on an as-and-when basis.
- **The 'Operationalisation' Approach:** In some ways similar to the 'on demand' approach, in this model again analytics professionals were primarily involved with client interactions and quantitative analyses, however, in this case the implementation of results was primarily into enterprise software. As such, the decision support aspects (outputs) were operationalised, and this required involvement of the technical teams that managed and developed these tools.
- **The 'Technical + Business' Approach:** The final approach differed from many of the others in that front-end activities were managed separately from 'technical' activities. In other words, there were teams of 'business analysts' who managed the softer aspects of each project (such as problem structuring or presentation of results), and technical teams, which in this case required the composite of technology-based skills (such as programming and database management) and quantitative skills.

8.1.5 Interviews with Academics in Analytics

The final instrument, presented chapter seven. was again based on interviews, this time with those involved in teaching and developing analytics degrees in universities in the UK and Ireland. Again, a template analysis was employed, utilising the structures developed in chapter six.

This analysis had two main goals. Firstly, its sought to complement chapter five towards meeting research objective five, identifying the skills taught within analytics degree curricula. Secondly, it was designed to meet objective five, identifying the potential barriers and the concerns which may influence how degree curricula are developed. The former of these, which requires some synthesis of results, will be addressed in section 8.3 of this chapter. However, some summary of the identified barriers is given here.

One of the most immediate issues is the sheer volume of content that may be associated with analytics courses. This was something already visible from the earlier quantitative analysis of chapter five (as well as the analyses associated with employer requirements in chapters four and six). However, from the perspective of course designers in analytics, it is effectively impossible to cover everything within a one-year course at the requisite depth for master's level courses. This necessitates some flexibility in our recommendations, or at least a recognition that there may need to be more than one type of analytics course.

Another cited issue, which has some overlap, was the available skills for such courses. In more than one interview, some desire to offer new content was suggested, but was prevented by a lack of expertise in the relevant field. One of the possible 'fixes' to such a situation, utilising skills elsewhere in the university (in other schools or faculties), was considered difficult (although not necessarily impossible) in many institutions, particularly if there was a disparity in tuition fees charged by different schools.

Thirdly, an issue that had not been forecast prior to the interviews (i.e. was emergent), more than one participant pointed to a potential lack of awareness of analytics amongst students. One reason for why this had not been identified *a priori*, according to one participant, is the nature of being in something of an echo chamber around analytics. For those within the field, or reading more general business literature (academic or otherwise), analytics, big data and data science seem almost overly-hyped and skills shortages (and the opportunities they present) frequently discussed. However, this does not mean such messages necessarily trickle-down to students, and more than one participant flagged the importance of ensuring potential, technically-minded students understand the opportunities that analytics may present.

Finally, there may challenges associated with the nature of academia itself in comparison to the nature of skills development. In other words, universities are not just training centres, and have other concerns and responsibilities beyond purely preparing students with a set of skills required by industry. Firstly, it was suggested that there may be an implicit trade-off in being a successful academic in this space (or, indeed most others) who has both teaching and research responsibilities. Teaching a degree in analytics requires (collectively if not individually) a wide variety of expertise, as already identified. However, research success is made easier by developing a very deep expertise in a particular area. Additionally, one participant flagged a potential disparity between the goals for analytics-type projects in academia, from those in business; the former being about generating theory and the latter about predictive power. In other words, organisations will care less about why something works, their concern will be that the models or algorithms do work and can create some value.

In the same vein, participants were also asked the extent to which universities *should* base their curricula on the perceived needs of employers, versus concerns related to the discipline itself. Most considered this as something of a balancing act. Ultimately, the inference from several interviews is that moreover there is a duty of care in essence to the students, to provide them with an appropriate set of skills to prepare them for their future careers. However, this is not necessarily just the immediate role they may assume after the degree, so they should be prepared for the long term not just the short term. Also, for one interviewee at least, this will not necessarily be what businesses, or the students themselves, think they will need, it is for course designers to create courses that feel best for the students.

8.2 The Skills Requirements of Analytics Employers

As discussed, meeting research objective three, defining the skills requirements for analytics professionals, requires the synthesis of the results of chapter four and chapter six (summarised in sections 8.1.2 and 8.1.4 respectively). Such skills are obviously diverse. Rather than produce one long, granular list, which to some extent has been done already in chapters four and six, this section will seek to address the issue in two ways.

Firstly, we will consider the main categories of skills and the most common within them (based around some of the topics of the template analysis of chapter six, but using insights from chapter four as well). Secondly, the chapter will reflect on the extent to which such skills are relevant to all flavours of analytics role, and if there are sub-groups within this which have different requirements. Through these discussions, research objective three will be met.

8.2.1 Skill Groupings for Analytics Roles

In the template analysis of chapter six, several 'nodes' of grouped themes from the interviews were identified (figure 29, section 6.2). Many of these represent key groupings of skills from which we can create the sort of higher level summary detailed in the introduction to this section. However, not all are necessarily related to skills (as the interviews covered a wider range of topics. The nodes excluded were "education", "analytics trends", "outputs" and "internal organisation", all of which will be addressed later in this section. As stated, to complete the analysis the results of this chapter (six) is also synthesised with those of chapter four.

8.2.1.1 Data and Data Management

One immediate topic of concern in this category is big data. In the interviews, whilst a few respondents reported they were regularly using what would be considered big data, for many the demand was primarily for 'small', mostly structured data sources. The relative lack of prominence, compared to what the 'hype' would suggest, was supported by the findings of the topic model analysis (chapter four), with the 'big data' topic only the 21st most frequent topic in adverts associated with analytics and those associated with OR (from the 56 relevant skills topics), behind less commonly discussed topics such as 'systems management' and 'customer management'. However, it is important to note that these analyses represent a snapshot of a moment in time, one which may already have become dated. Many interview respondents reported that they were beginning new initiatives in this area, seeking to hire staff with these skills, or at least "listening in", and the recruitment consultants both reported some increase in demand.

In general, however, database and data management skills were seen by many interview participants as very important to their functions, one even describing recruiting two staff members with database skills to their team creating "a revolution" (Analytics Manager, Telecoms) in how they worked (section 6.3.1). Many also highlighted the need for a coherent data strategy, as well as increased responsibilities for data extraction and data cleaning. This was not as obviously replicated in the topic model analysis, with the "databases" topic ranking only 24th overall for analytics job adverts, although SQL ranked as the most frequently requested skill in analytics job adverts (table 20, section 2.1.1).

In summary, some aptitude and experience with data and data management would appear to be both a key skill for analytics roles, and also one which we can reasonably assume to have greater importance than it may have typically attributed in OR-style courses. This latter line of thought will be continued in section 8.3.

8.2.1.2 Quantitative Methods

Quantitative methods, of course, were a frequent topic in both analyses (chapters four and six). OR methods were particularly prominent in the interviews, although, as identified in section 6.3.2, there was a high proportion of respondents who had some OR association, and therefore potentially not a completely bias-free sample. Whilst there was no dedicated 'OR' topic generated in the Online LDA of chapter four, there was, however, some evidence of its use in analytics job adverts. Firstly, 'analysis (quantitative)' was the fifth most prominent topic in analytics job adverts, as it was for OR adverts (table 19, section 2.1). Secondly, and more generally, analytics and OR job adverts showed relatively high correlation in respect to 'hard' skills (a mix of technological and quantitative skills), at 0.85 (table 21, section 2.2). However, it is worth noting that OR ranks only third for overall correlation (behind statistics and machine learning) and the 'modelling' topic ranked only 11th for analytics job adverts (compared to 2nd for OR jobs). Drawing from this, and other background literature (see sections 1.3 and 1.4 for examples), it seems sensible to conclude that OR methods have a role in analytics jobs, but that they are not the only show in town, and there is not a perfect overlap with traditional OR roles.

Data mining was also considered an important skillset in most interviews, although in a few, those who were tasked with more ad-hoc analyses, it was rarely used. In some cases, this had been an area of recent growth, in tandem perhaps with the growing reliance on databases and data management inferred in the last section, providing an avenue for identifying new projects or new potential value for the organisation. Again, there was no single topic identified in the Online LDA that directly mapped to data mining, although 'business intelligence' featured as the 23rd most prominent topics in analytics job adverts. Machine learning seemed somewhat under-represented in our interview sample (in comparison to the literature). However, there are some respondents who are extensively using these techniques. Machine learning is also surprisingly low in its ranking for analytics job adverts, at 31st position. In contrast, for operational research job adverts it is the 11th most prominent topic.

The final 'child' node of the *quantitative methods* code in the template was statistics. This was the most widely cited of all skill groups in the interviews. The obvious reading, that it is a very important skill, probably holds true, but it is also worth noting that statistics plays a key part in both machine learning/data mining and in OR, so therefore organisations which primarily use OR and little machine learning or data mining, would likely require statistics skills, but so too would those where the reverse was true. Again, there was no specific statistics topic in the Online LDA model, but 'analysis (quantitative)' ranked as the 5th most prominent topic in the analytics job adverts, and 'modelling' as the 11th.

8.2.1.3 Programming and Software

As described in section 6.3.3, the *programming & software* topic in the interview template comprised of four child nodes. The first of these, 'general purpose languages' described the higher-level languages. VBA and Java were the most frequently cited, followed by the C family (C, C++, C#). 'Big data languages' (such as Pig), the second child node, were markedly less frequent, and in fact many considered there to be more of a move to interfacing to them from other languages than using them directly (e.g. wrapping MapReduce commands for Hadoop into SQL commands). The third node, 'statistical languages and software', were by far the most widely used. Of these SAS was the most frequent, followed by R. with some mention of SPSS and other bespoke tools (particularly around simulation and optimisation). Finally, query languages, specifically SQL, was widely used and sought after.

The results of the job advert analysis were broadly in line with these results, albeit with a few exceptions. Although prominent in the interviews, VBA did not feature greatly in the job adverts evaluated. This may possibly be due, or due in part, to its integration with Microsoft Office. Potentially adverts may ask for "advanced Excel skills" or similar, rather than explicitly mentioning VBA. On the other hand, many of the web languages (HTML, CSS and JavaScript) were more prominent in the job adverts, analytics roles in particular, than in the interviews.

There was some disagreement as to the extent that programming is a necessary skill for recruits and graduates to already possess. For some interviewees, such skills were pivotal and sometimes lacking in UK graduates; for others, these were skills that could be learnt "on the job". Another perspective given, is that general purpose languages such as Java were more for those who were integrating analytics into business systems, while statistical languages would be used for the initial analyses and model building. Seemingly this was particularly the case for those using the 'on demand' and, even more so, 'operationalisation' approaches.

8.2.1.4 Soft Skills

Soft skills are both a key concern for most interviewees, and also an area where some perceived there to be some gap between what skills are required, and what graduates typically may come in with. As detailed in section 6.3.4, this topic had several child nodes in the template. To summarise though, we can position the main concerns as pre- and post-modelling activities.

For the former, many highlighted a need for employees to be able to understand the problem or business needs. This also entails being able to ellicit requirements and information from stakeholders, and also understand and, to some degree, empathise with the 'customer' (whether that be a literal customer in the case of consultancies, or an internal contact for in-house analysts). None necessitates a specific approach to these tasks, such as employing the soft systems methodology, but many considered an awareness of these approaches would be a definite plus. In the post-modelling stage, almost all respondents stressed the importance of soft skills in communicating the results to 'customers'. A variety of methods were used for this, from workshops to informal discussion, but each of these makes some demand of staff to effectively communicate. Indeed, many argued that this of growing importance.

In the job advert analysis, soft skills were equally prevalent. 'Communication skills' ranked as the 3rd most likely topic in analytics job adverts, with 'sales skills' (7th) and 'consulting' (9th) also in the top 10. Although slightly less obviously linked, 'management skills', which was classed as soft skills (as opposed to hard), was ranked 1st of all topics. OR and analytics jobs show high correlation in this regard, at 0.849 (section 2.2).

8.2.1.5 Other Skills

Finally, for this section, a summary of the category of 'other skills' is presented. The most tangible of these is project management, a skill many thought useful, though few thought to be critical in a formal sense. Some followed variants of popular methodologies (Prince2 and agile were both cited), but for many this was more informal, but nevertheless impactful on project outcomes. Project management also had its place in the job advert analysis, ranked as the 14th most likely skill topic in analytics jobs.

Another important aspect from this part of the template analysis was domain experience. This was, however, something many felt the right candidate could develop over time. Many domains were also prominent in the job advert analysis, and interestingly it was this category of skills which OR recorded the lowest level of correlation (0.752, and its lowest discipline rank behind statistics, machine learning and computer science).

Finally, some respondents expressed a desire for candidates to engage with some aspects of business strategy, particularly in respect to how data was managed and organised in their organisation, and there was some talk of the importance of a certain mindset for the work. The latter was described as the combination of problem solving, problem structuring and independent and critical thinking.

8.2.2 Groupings of Skill Requirements

Whilst there are clear patterns in terms of responses, it seems overly reductionist to think that these can be condensed into a ‘one-size-fits-all’ list. In many cases there were multiple schools of thought regarding skill requirements across the interviews, some of which were touched upon in the previous section. For instance, whilst most respondents primarily worked with traditional data sources, in at least two of the interviews the implication was of big data sources being utilised more than others. Similarly, there were some respondents who regularly used data mining and machine learning, and others who rarely did; some respondents who argued that programming was a key and missing component, and others who argued the very opposite. Whilst some generic requirements can be offered that should be considered of reasonable importance in all analytics degrees, there is implication that some variety would be desirable to meet different types of roles that graduates may take.

To some extent a categorisation has already been made, in respect to the six approaches detailed in section 6.3.9 (and re-presented in 8.1.4). However, whilst these are useful, and certainly inform upon a categorisation of skills, they are ultimately designed to describe just one facet of the roles, or, moreover, the organisational structure within which they are situated.

Additionally, we may consider the results of chapter four, and the comparison between the topic distributions associated with analytics job adverts and those of OR (and other disciplines). Consistently, the three main disciplines in terms of correlation with analytics, were statistics, machine learning and OR. However, where the correlation occurs shows some variety. Machine learning has its highest correlation with analytics in terms of soft skills and domains, with hard skills its lowest (albeit still correlating at 0.721 and the third closest of the six disciplines). By contrast, OR correlates strongest with hard and soft skills, but has the lowest correlation in terms of domains (behind both statistics and machine learning, but also computer science). In effect, this tells us something of the 'average' of analytics, suggesting a greater alignment to IT-affiliated domains than would be expected for OR (web and digital is an obvious example from the chapter), but also would, at least on the evidence of this dataset, have closer alignment to OR methods than to machine learning methods. This is a result broadly supported in the interviews.

However, it may also say something about the variation. On the basis that we both observe these differences, but at the same time maintain high correlation even in the least-aligned areas for each discipline (effectively all correlations are above 0.72), this may be read as evidence of there being some aspect of analytics that is closer to machine learning and computer science, but another aspect that is closer to OR and possibly information systems. In essence, we may consider there to be some analytics roles that could be considered IT roles, and others that are better described as business-facing roles. This again is supported by discrepancies between interviewees regarding the importance of programming, the importance of databases and big data, as well as the use of data mining and machine learning.

This also has some resonance with the five approaches listed in section 6.4.7, particularly as summarised in figure 40 (for convenience, shown again here as figure 41). Again, the idea is that there is some degree of a range of values, between more IT-orientated roles likely to involve more integration with the *technological* aspects of analytics and an emphasis on machine learning or data mining and business intelligence, to a capacity that more closely resembles the OR tradition, typically where more emphasis on integrating with the business process and/or function may made. In other words, the former may have more of a technology-basis, and the latter more of a 'people' basis. However, as figure 41 indicates, this is not necessarily an either/or relationship, with some companies at one of these extremes, but others more central, and therefore combining both a technology and a people focus in their analytics teams.

Figure 41 The five approaches in relation to role type (IT or business facing)

8.2.3 Summary

This section, whilst not specifically detailing a prescriptive checklist of required skills, has synthesised some of the results of chapters four (topic model of job adverts) and six (interviews with employers), and sought to find some of the communalities and contrasts, and reached some summary, albeit one which has various sub-groupings within it. In doing so, this effectively combines and concludes the contributions of those chapters towards addressing research objective three, identifying the skills requirements for analytics roles. To complement this, and performing a similar task in the sense that it synthesises the results of chapters five and seven, the next section will seek to conclude the fifth objective, identifying the current provision of analytics courses in UK universities.

8.3 The Analytics Syllabi of UK Universities

As the last section sought to summarise the requirements of employers, and synthesise the results of the two instruments designed to address this objective, this section will seek to do the equivalent for our analyses of the current provisions of UK universities for analytics. In doing so, objective five will be addressed, identifying the content, subjects and applications currently taught in analytics and OR degrees. The two instruments which are used to do this, are the machine learning model of online course materials presented in chapter five, and the interviews conducted with academics working in analytics presented in chapter eight. The section will be comprised of two components. Firstly, we consider the contents of OR degrees and related fields, and secondly the same will be done for analytics degrees.

8.3.1 The Curricula of OR and Related Degrees

As part of the analyses of chapter five, based on a bagged support vector classifier (SVC) and moreover the module analysis (section 5.2), some characteristics of OR degrees were identified, as well as for other related disciplines. Most notably, the Module Term Weighting (MTW) score was calculated to evaluate the modules which had the highest discriminatory power. In other words, borrowing from the χ^2 test, a score is given to modules such that a module that occurs frequently in a specific discipline, but is infrequent in other disciplines, it will score highest; while modules that are frequent across all disciplines (such as ‘research methods’) or are infrequent in the discipline, will score low (see 2.5.2 for a full description of the scoring system). Using this, figure 22 (section 5.2) was produced, which effectively provides a summary of different degree types and the modules that are most unique to them. For convenience, this is repeated below as figure 42.

Computer Science	
Term	**MTW**
Graphics	0.5406
Distributed computing	0.5224
Computer architecture	0.5197
Mobile	0.5057
Internet programming	0.4947
Software	0.4649
Computer security	0.4592
Programming	0.4531
Multimedia	0.4304
Networks & servers	0.4249

Information Systems	
Term	**MTW**
Strategy	0.8140
Performance management	0.7821
Enterprise resource planning	0.7804
Management	0.7764
Information systems	0.7415
Knowledge management	0.7094
Project management	0.6897
Business intelligence	0.6632
Operations management	0.6486
Human resources	0.6486

Machine Learning	
Term	**MTW**
Robotics	0.8416
Natural language processing	0.8130
Image processing	0.8045
Machine learning	0.8026
Computer vision	0.7691
Visualisation	0.7581
Artificial intelligence	0.7104
Business intelligence	0.6627
Agents	0.6607
Neuro science	0.5777

Operational Research	
Term	**MTW**
Spreadsheets	0.9420
Supply chain management	0.9171
Operational research	0.9137
Decision sciences	0.9033
Operations management	0.9033
Game theory	0.9033
Optimisation	0.8943
Consulting	0.8937
Simulation	0.8922
Stochastic modelling	0.6713

Psychology*	
Term	**MTW**
Psychology	0.8954
Business psychology	0.8954
Social psychology	0.8954
Cognitive psychology	0.8954
Clinical psychology	0.8954
Neuro science	0.8431
Human resources	0.7211

Statistics	
Term	**MTW**
Bayesian statistics	0.8701
Hierachical data	0.8701
Experiments	0.8701
Surveys & sampling	0.8701
Linear models	0.8701
Regression	0.8701
Survival analysis	0.8701
Geospatial	0.8701
Monte Carlo	0.8701
Medical & health	0.8588

Figure 42 Top 10 modules per discipline based on module term weighting

Analysing this data for OR, the list of modules is perhaps familiar to those working on, or studying for, OR degrees. The order may be surprising, but it is worth remembering that this is based on 'uniqueness' not frequency. 'Optimisation' and 'simulation' both feature in the top 10, but lower down the list than may be expected if it were in order of frequency, suggesting they are still common, but also feature in other disciplines (optimisation is of course a significant topic in machine learning, and simulation in statistics).

The highest scoring is 'spreadsheets', with typical module titles such as "spreadsheet modelling". This echoes the desirability and/or requirements for VBA skills suggested section 8.2.1.3. However, it does also suggest some difference from other degrees where modelling and data analysis are prominent, particularly statistics and machine learning. The inference would be that such degrees are unlikely to perform such modelling in Excel or other spreadsheets.

Other noteworthy inclusions are 'decision sciences', supporting our characterisation of OR as in the intersection of 'quantitative methods' and 'decision making' in the Venn diagram of figure 14 (section 2.3); 'consulting', another important skill identified in the interviews with employers (section 8.2.1.4); and other common OR applications in 'game theory' and 'stochastic modelling'.

Across the other disciplines, significant face-validity can be reached: with computer science focusing on programming and architecture; information systems including areas such as 'enterprise resource planning' and 'knowledge management'; and various statistical methods featuring for statistics. There are a few interesting results. One is that 'business intelligence' features, and with similar weighting, for both information systems and machine learning, suggesting this may be a point of overlap. Also, psychology and machine learning both feature 'neuro science', though it may be assumed that for the former this is more orientated towards an understanding of the workings of the brain, whereas the latter in replicating these methods algorithmically (i.e. neural networks and deep learning).

The interviews with academics, presented in chapter seven, was primarily focused on analytics courses. However, some inferences can be drawn about the nature of OR courses. In some cases (names not listed to maintain confidentiality), the universities offered courses in 'analytics', as well as in 'operational research' or 'management science'. For one respondent, the difference between the two was that the OR-labelled course had a greater mathematical content (and therefore pre-requisites for a quantitative background).

For others, both macro-level and micro-level differences were reported. For instance, one participant (coded as 'Professor, Ireland/Northern Ireland') argued that at a macro-level, analytics would be more focused on organisational strategy whereas OR on an operational and/or tactical level. In terms of methods, their view was that simulation had less relevance in analytics, compared to OR, with data mining being the opposite.

In summary, and to some extent synthesising with the conclusion of the previous section (8.2), there are inferences that OR degrees have some difference in both methods and in their ultimate application. OR, compared to analytics and on average, will have more focus on certain methods (such as simulation), certain software (such as spreadsheet software), and in certain domains (most notably areas such as supply chain). There is also some inference that OR will favour problems and application areas around the operations of an organisation than its overall strategic concerns.

8.3.2 The Curricula of Analytics Degrees

Having discussed some of the aspects of OR degrees in the previous section, this part will consider the contents of analytics degrees, again drawing on the analyses of chapters five and seven. However, as identified in section 5.3, and summarised above in section 8.1.3, the findings of the quantitative analysis of chapter five was that analytics degrees, in the main, fall into one of two 'types'; those most closely associated with machine learning (type one) and those associated with OR (type two).

The first type had a much stronger association with computing and technology, with the majority of these degrees based in schools or faculties baring these names. This is also suggested by the modules they include, with the five most likely modules as: 'big data' (9.46% likelihood); 'data mining' (5.99%); 'machine learning' (5.36%); 'web & eBusiness' (5.05%); and 'programming' (4.73%). In respect to the interviews with employers (chapter six), the implication is that such degrees are aligned to the 'technical + business' approach, where the *technological* and *quantitative* aspects of analytics projects are performed by the same individual or team, and the *decision making* aspects managed elsewhere.

There are some inferences that can be made from the interviews with academics (chapter seven) regarding these 'type one' degrees, but in the main the interviews were with those working on 'type two' degrees. This is particularly evident from evaluating table 34 in section 7.4, showing only 2 of the institutions in the sample had significant coverage of machine learning, both the classification given to 'type one' degrees and its 3rd most likely module. Indeed, in one of the interviews, for instance, and supporting the results of the classifier, one respondent (Head of Department, North West) argued machine learning "is more computer science really". There is a similar disparity between the teaching of big data, the most likely module in 'type one' degrees, but something not significantly covered in our sample of 'type two' degrees. Indeed, another responder (Associate Professor, Midlands), suggested such content is problematic as they "wouldn't have anyone who is an expert in Hadoop".

In regard to "type two" degrees, which the SVC classed as OR, the association was moreover with business schools. This can be seen in particular with the most likely modules, finance and marketing (7.20% and 5.46% respectively). Additionally, there was an emphasis on quantitative modules such as statistics (5.21%), data mining and data analysis (both 4.71%).

Obviously, such courses are seemingly better represented in the interview sample. Most respondents reported similar coverage of business subjects (often as electives), and of quantitative subjects. Overall, there are overlaps with OR courses present in the sample. Indeed, many highlighted this association, some arguing there was little or no difference between the two (analytics and OR). Irrespective of whether they regarded there to be a conceptual difference between the two, many identified some changes to their delivery in the "analytics age". Several emphasise there is more interdisciplinary content offered, particularly that which may be considered more closely linked to the technological aspects of analytics. Of these, data mining, information systems and new software solutions were most widely cited.

In summary, the analysis of chapter five has found evidence, as discussed, of two types of analytics degree offered. One group, which we may more illustratively call *data science* (the most common degree title), that incorporates machine learning and computing topics (amongst other analytics curricula). The other, which to distinguish from analytics as a whole, and to note some differences from OR, we may call *business analytics*, incorporates business topics, quantitative methods, and, of course, more of an OR influence.

However, as identified in this chapter, even with these 'type two', *business analytics* degrees, there is some adoption of new content and a move towards more computing and data-orientated topics. Indeed, as much as we find evidence for a two-tiered typology, there is seemingly not a pure dichotomy, with some interview respondents further towards 'type one', *data science* content than others. Therefore, it would be more illustrative to conceive this as a range of options between each type, which can also be extended to include the hard and soft skills discussed in analytics job adverts (chapter four) and in interviews with employers (chapter six). Following the logic, figure 43 summarises this perspective.

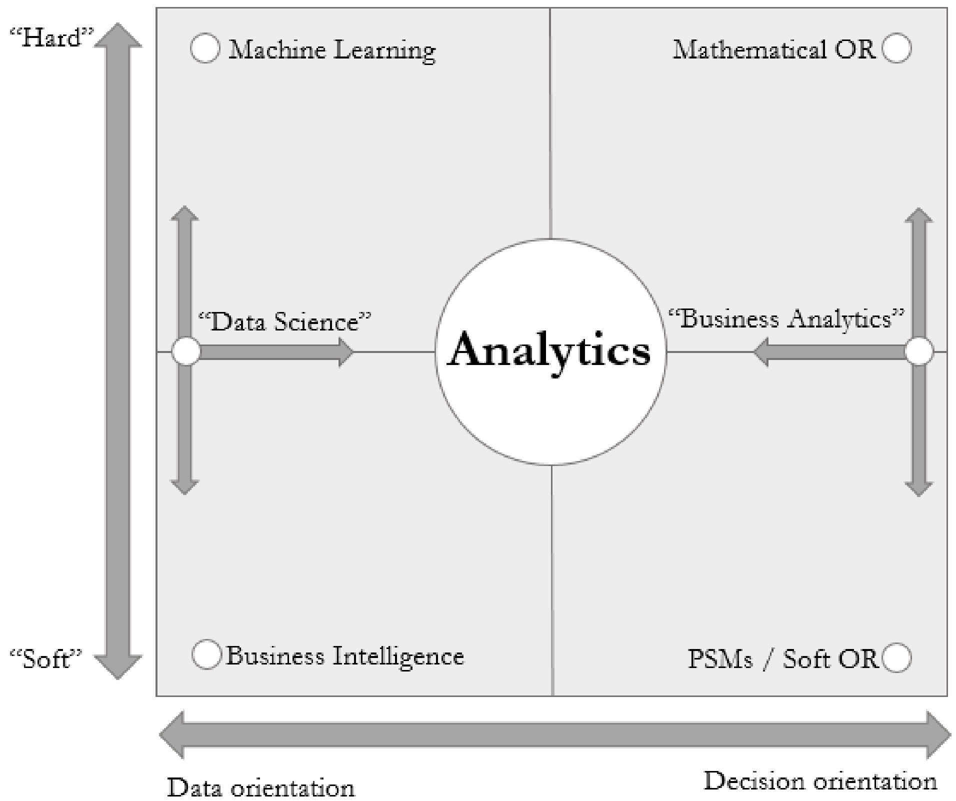


Figure 43 The content options for analytics degree curricula

In this conceptualisation (figure 43), the axes separate hard and soft incarnations of each discipline, and by their orientation (either towards data, as in *data science* courses, or to specific decisions and problems, in *business analytics* courses). Just as OR has incorporated both hard and softer aspects (from the purely mathematical to problem structuring methods (PSMs) and soft systems methodologies), analytics courses too may incorporate either or both (with the latter more closely meeting the perceived requirements of employers). At the data-orientation end of the spectrum, again we can argue for such distinctions, between the harder aspects of the cutting-edge of machine learning (such as neural networks and Bayesian processes), to the practices of business intelligence (which combine some technical aspects, but may necessitate developing a business understanding and on the visual presentation of metrics and data).

For a degree of either type, *data science* or *business analytics*, there is potential to take a variety of directions; such as the mixture of hard and soft skills offered, and the degree to which they also seek to include aspects of the alternative orientation (data or decision). Consequently, the final recommendations of this research, which will form the content of the subsequent and final section of the chapter, need to also consider three aspects: *data science* degrees; OR degrees themselves; and also *business analytics* degrees which will likely operate in some middle-ground between these two.

8.4 OR and *Data Science* Curricula

As detailed in the previous section, the final part of this chapter will seek to make the recommendations that can address the final objective of this research, creating a framework for the development of analytics and OR degrees, as the sum of the insights from each of the research instruments presented. However, the analysis of current academic provisions from the previous section also highlighted the need for this to form a three-pronged approach; covering *data science* degrees, OR degrees, and *business analytics* and *analytics* degrees. As the last of these is considered the principal output of the research, it will be considered in the next section of the chapter. However, this section will consider the former two, beginning with *data science* degrees.

8.4.1 *Data Science* Degrees

The first of these is the 'type one' degrees of the quantitative analysis of chapter five, labelled here as *data science* degrees (as the most common degree title). By default, these degrees are the furthest from the OR discipline, thematically. It is, therefore, the hardest for the OR discipline to reach and to influence, and to some extent beyond the reach of these recommendations.

However, it is worth recalling some of the insights regarding the OR research community and analytics; that to some extent there is a competition for customers (section 3.3). Whereas in that case the competition was for potential business users of different analytics technique, here the customers are potential students. Additionally, many of the academic interviews highlighted a potential awareness gap amongst such students of the potential value and opportunities analytics degrees may bring (section 7.7). In combination, these factors mean such degrees remaining a concern of the research.

In this spirit, there are two recommendations that can be made. Firstly, there is a more combative direction, competing with *data science* degrees. This would work by promoting the value of the OR and *business analytics* approach, to students and to potential employers. While this has obvious value, and something groups such as the OR Society already engage in (for instance, the 'O.R. in Schools' initiative, https://www.theorsociety.com/Pages/ORinSchools/ORinSchools.aspx), it is unlikely to be the sole solution as resources of this kind are likely to be limited and are already employed in this way.

Secondly, there is also a more collaborative approach possible. Provision of content within such *data science* degrees, or making resources available to students of these courses, can provide awareness of these methods to those likely to become future analytics professionals, even if that would be those working in the sort of 'technical + business' roles described in sections 6.3.9 and 8.2.2. Particularly considering that many consider that the likelihood will be that future careers will be increasingly dependent upon learning new skills (for instance, Smith and Meaney, 2016), increasing awareness amongst such a group has obvious potential benefit. However, whilst this line of thought has appeal, it may be naïve to consider this an easy option. In recent decades, many have observed an increasing commercialisation and corporatisation in UK universities (e.g. Robertson, 2010). With courses increasingly expected to show commercial value as well as academic, competition between departments may be the more likely response than collaboration.

8.4.2 OR Degrees

OR degrees are far from numerous in the UK; only 10 degrees with this (or similar) as a title were identified in the data collection for chapter five, compared with 43 analytics degrees and 69 computer science degrees. Despite this, there is evidence that OR degrees have value for employers. Indeed, many employers stated a preference for OR graduates, and a recruitment consultant in the analytics field stated they would be keen to engage more candidates from this background (section 6.3.2). However, with few degrees offered, and evidence of some being redeveloped, or perhaps simply relabelled, as analytics degrees (section 7.3.2), it is unclear the extent to which they will remain distinct from the *business analytics* degrees described in this chapter. In this context, extensive recommendations may be problematised, and we may moreover favour an 'inside-out' rather than an 'outside-in' approach; that recommendations for the role of OR working within analytics and analytics degrees, may have more impact, and more potential customers, than can be reached working as a discipline outside of analytics.

What is of value, however, both in the context of the curricula of OR and *business analytics* degrees, is to highlight the aspects of the OR tradition seen as most valuable by employers. The most obvious point is the quantitative techniques themselves. Many of the interview sample reported the use of OR approaches, particularly optimisation and simulation. Additionally, statistical approaches were widely cited as key, and in all the interviews modules on statistics were included.

In respect to 'softer' skills, OR too makes a valued contribution. Whilst few employers had a formalised use of soft system methodologies or problem structuring methods, many spoke of their value. Moreover, soft skills in general, particularly those related to communication and consulting, were highly prized by almost all employers. In this regard, OR has advantages over 'competitor' disciplines. The evidence of chapter four, particularly the strong correlation between soft skill requirements of analytics and OR job adverts shown in figure 19, and the similar distribution of skills (hard, soft and domain) shared by roles in each shown in figure 20 (both in section 4.2), would suggest that OR teachings remain very relevant to analytics professionals.

Finally, and whilst perhaps the most nebulous and intangible aspect of this discussion, nevertheless one which was reported as critical in many interviews (section 6.3.5), is the development of an *analytics mindset*. Although an argument was made that to some extent this was innate rather than necessarily something that can be taught, there are aspects which employers felt could be developed such as an aptitude for problem solving, abilities around the structuring of problems, and critical thinking. These aspects have long been acknowledged and discussed in the OR research community (e.g. Kaplan, 2008), and some of the interviewees specifically referenced OR degrees as helpful in developing these skills.

In all, clear argument can be made for the importance of the skills associated with the OR discipline. However, as stated, there is some question marks as to whether these will continue to be taught within specific OR-titled degrees, or as part of an *analytics* or *business analytics* curricula which maintain the influences of the OR tradition. The recommended contents of such courses form the discussion of the next section

8.5 *Business Analytics* Degrees: A Developmental Framework

The final area of discussion, and the most important to the argument of this work, are the analytics degrees most closely aligned to the OR tradition, those described as *business analytics* degrees (as opposed to the *data science* degrees discussed in section 8.4.1). These were the most numerous amongst the analytics degrees analysed in chapter five (20 out of 43 and exactly double the number of OR degrees identified), and were highly represented in the interview sample.

The recommendations for these degrees, will represent the main contribution of this chapter, as well as meeting the seventh and final objective of the research. To do so, the recommendations are presented in the form of a framework of topics, informed by the insights of all the previous analyses. It is worth, however, reflecting on how such a framework may compare to those identified in the prior literature, particularly the work of Lunt *et al.* (2008) into the design of undergraduate IT curricula (detailed in section 1.6.2.2). As highlighted in that section, there is some difference between their endeavours, and the appropriate framework for this case. Most notably is the degree with which the recommendations should form a prescriptive list.

For analytics degrees, we have already highlighted that this is something of a broad church. Not only are there more than one 'type' of analytics master's degree, but also, we have seen that there are multiple approaches and structures that are utilised by analytics employers, each of which bring different skills requirements. There will also be some concern about pre-requisite skills and the module structuring of master's degrees. As opposed to the standardised three-semester structure of undergraduate degrees, master's education has more variety meaning some students will study more modules than others. Additionally, the interviews with those who develop such courses highlighted some potential barriers to developing a uniform curriculum, most notably the availability of resources.

The implications of this, is that the recommendations presented here need to have some flexibility to allow for the tailoring of course contents to particular use-cases (e.g. job roles and student groups) and to suit the specialisations of the teaching staff. In other words, the recommendations will be presented at a higher-level of abstraction than in in the report on IT curriculum; distilling the areas of greatest importance, as inferred for the previous analyses, rather than at the granularity of particular techniques and domains. In this spirit, and borrowing from the visual representation of Lunt *et al.* (2008, p19), figure 44 presents the key elements recommended, which will form the basis of the remainder of this section.

Figure 44 The 'pillars' of the proposed analytics curriculum

8.5.1 Analytics Foundations

Perhaps more a necessity of successfully delivering a course of this kind as it would be a core topic, there will be some foundational material required in any analytics degree, including the *business analytics* degrees described here. Essentially, ‘analytics foundations’ is used here to describe the more basic skills and toolkits needed to complete more advanced materials later in the course. Precisely what this would contain would be dependent on several factors. Firstly, the overall approach of the degree. For instance, if there is a decision to use a particular software or programming language, there will need to be some introduction to it towards the beginning of the course. An additional consideration are any pre-requisites that form a part of the entry requirements of the course. If, for example, there is a pre-requisite that students have proven mathematical training (such as a quantitative first degree), there may be less requirements to cover more basic topics in mathematics and/or statistics. Finally, course designers may wish to include more of a theoretical or background introduction to analytics and its use cases. However, the primary goal of such material would be to introduce students to the toolkits, particularly in respect to the technical foundations, required to complete the later modules.

8.5.2 Data Management

In this context, ‘data management’ is used to describe the range of topics relating to the procurement, storage and pre-processing of the data that is necessary for analytics tasks. Arguably, this is one area where OR courses have traditionally given limited coverage, and therefore a departure from a ‘pure’ OR degree. This is represented in the ‘options’ presented in figure 43. As OR is positioned at the ‘decision orientation’ end of this axis, rather than the ‘data orientation’, it follows that an analytics degree of any flavour, including *business analytics* degrees, is likely to have greater focus on data, and therefore likely to include issues of data management.

Although in the interviews of chapter six, there was no complete consensus on the extent to which this is an absolute necessity for analytics roles, most respondents stated some requirements in this space. Indeed, one respondent (Analytics Consultant, Finance) expressed their belief that without a foundation in these areas candidates “wouldn’t stand a chance [in the jobs market]”. Another (Analytics Manager, Telecoms), spoke of a “revolution” being brought about by recruiting staff with database skills.

Another common theme in the interviews, as it is in the literature, was the importance of data cleaning and an exposure to ‘real’ (i.e. messy) datasets. Several interviewees criticised universities for not using such sources in their exercises, although in every interview with academics some use of ‘real’ data was included in the degree. However, as identified in not only some of the interviews with academics but also one of the potential employer interviews, to have students perform extensive data cleaning on every activity would be very time consuming and tedious.

In keeping with the stated aims of this framework, to focus on a higher-level abstraction of requirements, no precise techniques need to be stated here, but some of the topics that may be relevant, but by no means an extensive list, would include:

- Open data and web data (such as application programming interfaces – APIs);
- Data cleaning and merging datasets;
- Data architecture (e.g. warehousing);
- Databases (relational and/or NoSQL);
- Data dictionaries and documentation.

8.5.3 Data Analysis

The third element, and second 'pillar', and in many ways the natural follow-up to issues of data management, is data analysis. For the purposes of this framework, data analysis is considered as covering a range of topics including, but not limited to:

- Design of experiments;
- Transforming data;
- Data mining;
- Summarising data;
- Data validation.

Many of these were considered important topics in both the job advert analysis of chapter four, and the interviews with potential employers of chapter six. Design of experiments (DoE), was flagged as a critical skill in multiple interviews. Although, strictly speaking, not best described as purely the analysis of data, as it is in part at least concerned with the collection of data, it has been assigned to this category as skills-wise it is more akin to data analysis than data management. Ultimately it is based upon statistical knowledge, and an understanding of the analysis method rather than the storage or processing method. Also, of course, DoE has implications on how an analysis is performed, and therefore should be planned in combination.

Particularly with the growth of big data, particularly in its varieties (text, image, video and so on), there is even greater need for skills around transforming and understanding of data. While big data as a topic was not as frequent, in both job advert analysis and interviews, as may have been expected from a literature review summary, the number of interview participants who were either experimenting in this space or were planning to in the future, would suggest that this will become a more important topic over the careers of such graduates. Irrespectively, even with 'traditional' data sources there is need for ability to understand, summarise and process data for further analyses or modelling activities. This will also include many of the more foundational statistical techniques that may be considered a minimum requirement of an analytics professional, from descriptive statistics, measure of central tendency, variation and distribution, through to analyses such as the t-test, χ^2 test, ANOVA, and so on.

Although somewhat dependent on the company, and more specifically the data resources available to them, there was evidence for the role of data mining in the companies interviewed. In some cases, data mining had become one of the key activities in analytics projects. Additional evidence for this can be seen in the academic interviews (section 7.4), with all but one university including content in this area, with one participant (Analytics Professor, Northern Ireland/ Ireland) specifically naming data mining as a method that is more important to analytics than OR. This also is in-line with the argument of figure 43, that a move from 'traditional' OR degree content to analytics degree content necessitates a greater 'data orientation'.

Finally, we consider data validation. Again, some disambiguation is necessary, as it is possible to consider this a task of data management; namely a part of the process checking data for errors or unexpected values before importing into data warehouses (or similar). In this context, while not disputing the importance of such activities, the term is moreover used to describe a critical evaluation of data, particularly as to its veracity (or otherwise). As such, this makes some nod to the importance placed on critical thinking in the interviews with employers (section 6.3.5).

8.5.4 Modelling

The fourth element, and third 'pillar' on the initial foundations (figure 44), is modelling. An interesting insight from the job advert analysis of chapter four is the slight disparity suggested between the importance of modelling in analytics and OR roles. For OR, this was the 2nd most prominent topic, whereas in analytics adverts only the 11th. This can be interpreted as partly due to the greater 'data orientation' we are associating with analytics, as a 'decision orientation' all but necessitates some modelling of the problem situation and alternatives. However, it is worth noting that in the analytics job adverts there is likely a greater range of roles, from both data and decision orientations and also the space in between. Therefor we can assume modelling to be more important in some than others, and the ranking is only representative of the average. Also, the recommendations of this section are not to the extremes of a 'data orientation', as our characterisation of this space includes these as *data science* degrees (section 8.4.1).

Obviously, modelling is related to, and not entirely separable from, data analysis; particularly as we include the more basic statistical models such as the t-test in the latter. For clarity, modelling is used in this context to loosely describes analytical work that resembles the following processes:

- Forming a problem definition or goal of the analytics;
- Sourcing and preparing the necessary data to represent this problem/goal;
- Translating this to some form of mathematical and/or computer-based model (be that from the OR, machine learning, econometric or statistical traditions);
- The validation, verification and analysis of this model.

Such processes will be very familiar to an OR professional, but that is not to say that this 'pillar' is only about OR. In an analytics course, it is likely content of this kind will not only focus on OR methods and applications, but a slightly broader range of tools that may include econometrics, forecasting, predictive analytics, regression, statistics, machine learning and others. This would seemingly be more in line with some of the requirements of employers inferred in the prior analyses. Therefore, it is worthwhile presenting some explanation of each of these elements.

An ability to structure a problem was a key requirement of many of the interviewees. This may involve a variety of skills and activities, such as conducting interviews or workshops, as well as abilities in logic and critical thinking. As such, this element may require multiple learning methods, but one that seems both relevant, and commented on in both sets of interviews, would be proving realistic business cases for coursework, either within classroom settings or for student projects and internship activities.

Whilst sourcing and preparing data has been incorporated in the previous 'pillars', for modelling there may be extra requirements around processing the data to meet the analytical method being used. In many ways, this will be on a case-by-case basis, as different analytical approaches will all have different requirements.

In respect to building the final model, again there will be some variety dependent on the approach being employed. Clearly OR models would remain relevant in this space, however, there seemingly is demand for other approaches. Given the increased focus on data in the analytics space, it is unsurprising that many employers referred to statistical modelling, and many in the academic sample said that their analytics included more statistics than a traditional OR course might. Finally, although we have found closer associations between machine learning and the *data science* interpretation of analytics courses, there is a relatively strong case for its inclusion in *business analytics* courses. Specific content on machine learning modelling were relatively rare in our sample of academic interviewees, with only two institutions including any substantial focus on it. It was also somewhat mixed in our interviews with employers. In some cases, most obviously those employing the 'technical + business' approach (section 8.2.2), machine learning was widely used, but it was not by any means extensive across the sample. However, the hype associated with these methods of late, and the advantages afforded to such approaches in the modern era (most importantly the availability of data and of processing power), would suggest that it would be an area where continued growth and interest can be expected.

An additional concern, is how and where the models are built and implemented. In keeping with the non-prescriptive approach employed in this framework, no particular software or programming languages will be specified here, but it is of course a concern. On the basis that no analytics course will be able to teach every relevant technique, there will be clear benefit in enabling students to be able to source and implement new approaches based on online resources.

As described by one respondent in the academic sample (Emeritus Professor, North East), successfully implementing analytics projects does not require the complete mastery of a particular tool or programming language, as there are typically numerous resources online. Providing students with the necessary foundations in implementing some algorithms and models, along with the ability to source software libraries and code examples that they convert to their needs, would likely give them greater opportunities in the work place. Although the importance of programming was of some dispute in the interviews with employers (ranging from critically important to something they can be taught ‘on the job’), if a confidence and an ability to achieve to find and use such resources can be given to students, it has the potential to be an empowering process. Such a line of thought can be summarised in the well-known proverb: “give a man a fish, and you feed him for a day. Teach a man to fish, and you feed him for a lifetime.”

Finally, in this pillar we include the analysis, validation and verification of models. Again, this is very specific to the approach in hand, but an important aspect of analytics modelling is to ensure the quality of the model and the recommendations it generates.

8.5.5 Implementation

The final pillar concerns implementation. Again, a potentially ambiguous term, especially as it can mean very different things to the OR community and those working in software for instance. It is used here to describe the wide range of potential activities that are required to move from analysis and modelling, to the deployment of a model or it’s insights into the organisation.

On the basis that there was seen to be great variety in terms of the potential outputs of analytics within the organisations of the employer interview sample, it is recommended that analytics courses try to represent some of this diversity in their curricula. For some of the sample, the main requirement was the communication of results. Again, however, this can be in a variety of forms. Many cited written outputs, so best practice in generating documents and reports would be worth including. Presentations and workshops were frequently mentioned, requiring many of the soft skills around verbal communication and managing interactions with clients. Other companies sought to operationalise models into enterprise software, and one interviewee, Analytics Manager (Retail Travel), spoke of the importance of writing technical documentation for IT professionals. Finally, visual communication, most notably data visualisation, was used in many organisations, meaning familiarity with dashboard technologies or software such as Tableau and QlikView would be beneficial.

Given that analytics degrees will invariably introduce students to a variety of analytical methods, there is an opportunity for students to be exposed to many of these different implementation methods over these modules. In other words, if structured in the right way, degree courses can ask students to present their results in different ways in each of their modules. This can allow them to experience both something of the variety of analytical methods available, but also the variety of outputs and implementation methods.

8.5.6 Analytics Professionalism

The final aspect of figure 44 is analytics professionalism. Again, this directly borrows from Lunt *et al.* (2008), but is nevertheless both an important element, and one where distinct requirements for analytics graduates can be identified. In this context, analytics professionalism again covers a range of areas, focusing on the skills that are not necessarily a part of analytics itself (at least as it is defined in this research), but are a part of making analytics work in organisations.

An example of this is project management. Although few of the organisations in our employer interview sample rigorously employed a formalised project management methodology, many considered it an important skill. One respondent, Analytics Manager (Retail Travel), suggested that if they were developing an analytics course, they would include project management, but only as a relatively small part of an existing module. The recommendation here would be similar. Whilst it is probably overkill to make project management a compulsory, semester-long module, giving students a basic understanding of the key methodologies and of best practice seems a sensible approach.

Another suggestion in the employer interviews was to give graduates some understanding of how to structure analytics and analytics teams within an organisation. For many organisations, there are not only difficulties in recruiting analytics professionals, due to the forces of supply and demand, but also there are issues in organising these resources to maximise their effectiveness. Again, this unlikely to warrant a whole module, but seemingly is useful experience for students to bring to their future careers.

Additionally, a key and recurrent theme was in making problem sets and activities as realistic as possible, and providing 'messy' datasets and ambiguous scenarios (as opposed to highly artificial and overly simplified walk throughs of techniques). Whilst this was something all the participants in academic interviews said was included in their courses, seemingly this is an area in which there cannot be too much of a good thing. Collaborations with industry to obtain real case studies, real problems to try to solve, real datasets, and consultancy-style projects is highly recommended.

Finally, many interviewees mentioned the importance of an understanding or awareness of particular domains, and also certain domains were frequent in the job advert analysis. This becomes particularly relevant for this framework, because it was in respect to domains that the lowest correlation scores between analytics and OR job adverts are reported (section 4.2). The most likely domains for OR, perhaps unsurprisingly, included supply chain management, manufacturing and other operationally-led domains. By contrast, analytics job adverts more prominently featured domains around eBusiness and eCommerce, as well as a greater focus on marketing (which OR featured for, but not as highly). Focusing on the domains most relevant to modern analytics jobs would seem an obvious but important recommendation. This can be in the elective and optional modules offered, but even more immediately in the case studies, examples and problem sets used in classes.

8.5.7 Developing Analytics Courses

The discussion of this section has set out a framework, detailing of a set of recommended elements that may comprise a *business analytics* curriculum. These elements have drawn from the different research instruments employed, but designed to be relatively flexible and not unnecessarily prescriptive. Such flexibility is probably necessary, considering some of the limitations universities may face when developing such courses, such as the availability of staff to teach different elements. However, this flexibility may also be desirable, allowing universities to create differentiation and find niches within this space. This is particularly relevant in respect to the combination of analytics curricula with that of specific domains. For instance, with additional marketing modules, such a framework can be used to develop a MSc in Marketing Analytics, or with the relevant content, a MSc in Supply Chain Analytics (and so on).

Another note is that the pillars of these frameworks can be applied in multiple ways. For instance, an analytics course may include a specific module in data management or in data analysis. However, it may also seek to cover multiple or all of these elements in a single module, such as offering an optimisation module that covers the full process from data acquisition, to initial analysis, the optimisation approaches themselves, and then how they may be implemented into organisational practice.

Although flexibility has consciously been built in to allow some leeway for skills availability, that is not to say this means course designers may ignore important elements. For instance, data management may be something OR teams lack deep expertise in (alluded to in the interview with an Associate Professor, Midlands, albeit in the relatively extreme case of the Hadoop file system). The inferences from the interviews and job advert analysis is that this does represent an important set of skills for graduates and recruits, suggesting that efforts are necessary to ensure its inclusion. One solution is through greater collaboration with other departments. Though problematised by institutional structures and aspects such as fees, the increased diversity of topics in analytics (compared with OR), particularly in respect to the more computational and data orientated aspects of analytics, would clearly justify such efforts.

The recommendations and framework presented here are targeted at developing postgraduate level degree courses. However, to some extent there may be some overlap with the recommendations that could be made for the development of undergraduate curricula. In respect to the topics of figure 44, there may be little or no difference between the two, other than potentially the emphasis. For undergraduate courses, the aspects of “analytics foundations” will likely require more emphasis than at master’s level, where some assumptions on level of prior experience and training can be made, either implicit or via pre-requisites. Contrastingly, some aspects of “analytics professionalism” may be less emphasised at undergraduate level, as it may be reasonable to assume students to be less likely to take roles that require strategy and influence on the structuring of analytics teams in the immediate term.

One significant difference for undergraduate curricula is in their sheer volume; with courses lasting 3-4 years compared to the one year now typical for master’s degrees. This provides lots more opportunity to introduce specialisation into domains, as well as more diversity in general for the curricula. For instance, given these timeframes it is more than possible, and indeed recommended based on the comments of interviewees in the employer interviews, to include an extended work-based placement. It would also likely be possible to cover both the *business analytics* curricula discussed in this framework, along with what we might consider to be *data science* curricula. However, at the same time, it is worth remembering that there may too be significant difference in the students themselves, who may, on the averages at least, bring less maturity and prior experience, suggesting that the more technical aspects of the degree may have to be built towards more gradually and at a more introductory level.

8.6 Summary

This chapter has sought to perform three main tasks. Firstly, some summary of the previous chapters and analyses have been presented. Secondly, and in particularly in the discussions of sections 8.2 and 8.3 to address objectives three and five respectively, the findings of these separate analyses have been synthesised. Finally, the research has considered the curricula of analytics and OR degrees. For OR itself, and for *data science* type courses, general recommendations for the OR community have been provided. For the provision of *business analytics* courses, a framework has been presented, designed to accommodate the requirements inferred from the job advert analysis and interviews with employers, whilst also recognising some of the barriers and issues faced by analytics course developers. In doing so, the research has addressed the seventh and final objective of the research. To conclude this research and thesis, the final chapter will summarise the work, and highlight some limitations and opportunities for future research.

9 Conclusion, Limitations and Future Research

Over the course of the previous eight chapters, a variety of topics have been explored, research instruments employed, and results and insights presented. Through this process, all seven of the research objectives have been addressed, culminating in the framework presented in the previous chapter. To conclude the work, this chapter will summarise the main findings and contributions of the research, and give some discussion on the potential for further research options in this vein.

The chapter is arranged as follows. Firstly, a summary of the research objectives is presented, alongside how and where they were met, and their respective results. This is followed by some discussion of the potential limitations of the work, and finally suggestions for future research.

9.1 Research Objectives

As detailed in the introduction, this chapter will begin by summarising the objectives of this research and how they were met. Each will be discussed in sequence

9.1.1 Research Objective One

"To determine the relationship between academic definitions of analytics, operational research, and other related fields and disciplines".

The first objective, a necessary step in setting the context of the research and the terms with which its analyses are conducted, was addressed in chapter three. The key argument made, to this end, is that there is in fact a shared ecosystem within which OR, and other disciplines related to *technologies, quantitative methods* and *decision making* co-exist and, to some extent, compete for customers, end-users of their methods and tools in organisations. This ecosystem is characterised as the *dianoetic management paradigm* which has been the dominant management philosophy of the last 100 years, where decisions have been sought to be made based on analyses and discursive reasoning. Within this context, analytics is considered to be simply the latest incarnation of this paradigm, and as such a space shared between OR and the other disciplines discussed in the chapter (and elsewhere in the thesis).

9.1.2 Research Objective Two

"To develop a research agenda for the OR community which addresses the concerns associated with analytics".

Also addressed within chapter three, the result of this objective is based upon the argument that OR is a constituent part of this management paradigm, and therefore a part as well or analytics without being precisely the same. Two extreme responses to this situation were presented: the *isolationist approach* where OR ignores analytics or any other cycles within the paradigm, instead taking an insular attitude to the topics and contents it chooses to focus on. As alternative, the *faddist approach* would be to completely rebrand and reposition the discipline around the dominant concerns and trends within analytics. Both positions were shown to be problematic and undesirable, with a position somewhere between these extremes recommended.

However, despite the potential benefits the visibility which analytics currently has may bring to OR research, section 1.3.3 identified what is described as a "publishing paradox", as comparatively little OR research has been performed into analytics, and in select few venues where it has. In consideration of these two factors, chapter three recommends a (non-exhaustive) series of topics for the OR research community to investigate, that are both relevant to the traditions of OR, but also prevalent in the broader analytics literature. These topics are big data, new data architectures, unstructured data, real-time analytics, and data visualisation

9.1.3 Research Objective Three

"To determine the skills requirements of analytics roles and the extent to which these may be met by OR professionals".

Objective three was addressed using two of the research instruments presented, the job advert analysis of chapter four and the interviews with employers of chapter six. These results were synthesised in the previous chapter, section 8.2, where the objective was achieved. The results include three main findings. Firstly, a series of key skills and experiences were highlighted across a range of topics including data and data management; quantitative methods; programming and software; soft skills; and the outputs of analytics (and the skills they necessitate). Secondly, analyses were performed to compare the similarity of analytics and OR roles (chapter four). The main findings here were of an overall high correlation, particularly in relation to 'hard' and 'soft' skills, although greater differences were seen between the domains the two focus upon. Finally, in section 6.4, five different approaches to how analytics is structured within the organisations of the interview sample were presented, with discussion on how this impacts on skills requirements.

9.1.4 Reserarch Objectve Four

"To identify the academic disciplines with which analytics master's degrees most closely align".

Objective four was addressed using a bagged support vector machine, trained to classify analytics degree material as aligned to OR, computer science, information systems, machine learning, psychology or statistics materials (chapter five). Although some degrees were classified as either statistics or psychology, the majority (86%) were classed as OR (20 out of 43) or machine learning (17 out of 43). Based upon this insight, and also some analysis of the modules offered on analytics degrees, a typology of analytics degree was presented, with those most associated with machine learning, labelled *data science* in chapter eight; and those associated with OR, labelled *business analytics*.

9.1.5 Research Objective Five

"To identify the specific skills, subjects and techniques taught within analytics degree curricula".

Objective five was addressed by the synthesis of the results of chapter five and chapter seven, presented in section 8.3. Building upon the two approaches of the typology of chapter five, *data science* and *business analytics* courses, this was presented as two separate groupings, although obviously with a reasonable degree of overlap. *Data science* courses displayed closer association with computing or technology schools, featured IT-orientated content and moreover displayed a data orientation. *Business analytics* courses, on the other hand, had stronger association with business schools and a range of business topics, and what is described as a decision orientation; a focus more on specific problems or scenarios to model. In each case, there are some differences in necessary skills and experience, which are presented and discussed in chapters five and eight. Whilst these two categories of analytics degree show separate characteristics, it is also recognised that rather than pure, binary dichotomy, analytics degrees can offer varying degrees of each type, and a range of positions, also including variation between a focus on 'harder' or 'softer incarnations of each, was presented in chapter eight.

9.1.6 Research Objective Six

"To identify the potential barriers and concerns that impact on the creation of analytics and OR curricula".

The penultimate objective was addressed in chapter seven and concerned the issues, limitations and barriers faced by course designers in developing analytics master's degrees. Several such issues were identified, including availability of specific expertise; difficulties in collaborating across faculties or schools; incorporating the varied elements of analytics into a relatively short schedule of modules and classes; and, unexpectedly, a lack of awareness amongst potential students of what analytics is, and the opportunities it may offer graduates. Some suggestions for mitigating these are given in chapters seven and eight, and were considered in the design of the final framework of chapter eight.

9.1.7 Research Objective Seven

"To create a framework for the development of analytics and OR degrees".

As mentioned in the previous section, chapter eight also included a final framework for how analytics degrees may be developed and the key concerns their curriculum should incorporate, thus addressing objective seven. We postulate there to be three areas of concern, *data science* type degrees, OR degrees and *business analytics* type degrees. Some recommendations are made on how the OR community may respond to *data science* degrees, including both a recognition of the competition they may present, as well as the opportunities for collaboration and to reach new audiences they offer. OR degrees are briefly considered, shown to be still of value based on the findings of chapter six (in particular), but seemingly less common than analytics and even *business analytics* degrees.

However, it is towards the curricula of *business analytics* that the greatest contributions are considered possible, and a framework of recommendations is presented (again in chapter eight).

The framework consisted of six elements: analytics foundations; data management; data analysis; modelling; implementation; and analytics professionalism. Each of these was discussed in sequence, with tangible recommendations made. The framework is designed to be somewhat flexible, and to consider some of the issues and barriers identified in chapter seven (as well as incorporating the insights from each of the previous research instruments). Additionally, to this end, further recommendations are given as to how such courses can be designed and developed.

9.2 Limitations

This research has covered a wide range of issues and perspectives, towards the goal of reaching a better understanding of analytics, and the job roles and curricula requirements that may be associated with it. In presenting our findings, one aspect that may seem conspicuous in its absence is a sense of precision in respect to the specific tasks that an analytics professional may require expertise in, and a clear picture of their daily activities. Ultimately, our results and discussions speak more to an overarching set of practices and higher-level groupings of skills.

This may indeed be a limitation, and restrict the some practical application of our findings. However, in analysis this seems to be more appropriate to the picture of analytics that has emerged; a picture of a 'profession' that is more superset than specific, and seemingly affords a range of organisational structures, applies a variety of analytical methods, and can result in multiple different job specifications. In other words, "analytics" is a broad church, and can be an appropriate descriptor for a variety of job roles and backgrounds. What has been achieved in this process, however, is the identification of some common patterns and/or groupings, and some general recommendations as to how degree curricula can be developed to prepare graduates for such an environment.

Additionally, with multiple research instruments, there are too multiple sets of limitations. The historical analysis of chapter three (objectives one and two) was ultimately literature review based, and drawn from a specified set of source disciplines. Possible limitations are introduced by the choice of material and the scope of the literature survey.

Even within these selected sources, it is notable that the sources are used to explain the periods in a general sense, which does not necessarily account for the extent of diffusion across all, or even the majority of, businesses. Indeed, many businesses will exist entirely outside of the paradigm described, using little to no information technology or analytical approaches in decision making. In many ways, the analysis may be more 'histographical' than 'historical', as the periods are described based on the literature written in or about them, rather than empirically analysing artefacts or interviewing those who directly experienced the time periods. This is an obvious limitation in regard to the chapter measuring the periods as business practice, but at the same time retains value as a measurement of how the periods were described.

The job advert analysis of chapter four (partially meeting objective three) introduced limitations particularly around the data source. Firstly, being based on text data, there are issues and subjectivity introduced that is typical of analyses of this kind around the extraction and processing of the data. Also, the analysis relied upon the interpretation of the topics generated, which is ultimately a subjective task. Finally, it is important to note that the data is collected in a cross-sectional way, and therefore only representative of a snapshot in time.

Chapter five presented an analysis of online materials concerning analytics degrees, and those of related disciplines. Again, this was based on text data, with the issues this presents. Also, it is relevant that the model was actually trained on job advert data, as there was only limited course data available. This presents a major limitation, somewhat tempered by the fact the model performed better when classifying degree materials than job adverts. Irrespectively, the accuracy and F_1 scores were less than ideal, and future research may work on improving them. Finally, although appropriate to the scope of this research and necessary considering the time-consuming nature of manually extracting the data, the focus was only on degrees offered in UK universities.

The final instrument used is interviews with both potential employers of analytics and OR graduates, and with academics and university staff working on analytics programs. The most notable limitation was that there was some degree of bias in the sample towards the OR/*business analytics* interpretation of the field, in particular in the academic interviews. This is primarily due to access, and the fact that the participants in the academic interviews were recruited prior to identifying the two-class typology of chapter five.

In the face of the challenges inferred in the above, several mitigation strategies were employed. Most obviously, given that all of these instruments have their limitations, the mixed-methods approach of this research offers some mitigation in that the different instruments can be used to triangulate results, and to correct for biases introduced. Additionally, and relatedly, whilst there is the potential for error to be introduced via the data sources (particularly working with text data), by using multiple, independent sources this is somewhat mitigated. Finally, as several of ideas that are presented in the recommendations were formed using multiple instruments, they also were formed over a period of time. This has afforded the opportunity to present and test some key concepts in presentations and written papers, where external opinion has helped validate and shape the direction of the work.

9.3 Future Research

Future research around each of the individual instruments can be suggested, and again some of these are presented in the sequence they are introduced in this thesis. However, there are also possible future work in the direction of this research as a whole, which are presented at the conclusion of this section.

Firstly, the computational literature review (CLR) of analytics presented in chapter one could be contrasted to similar analyses of the OR cannon, or similar fields. While the nature of the CLR, as essentially latent Dirichlet allocation is a data-driven, dimension reduction approaches, it is unlikely that there would be value in a direct quantitative comparison between two separately generated topic models of the literature, a qualitative comparison could be of value. Alternatively, modified version of the CLR could be used, where the search keywords (in this case analytics and OR, could be included as co-variates in a combined analysis (see Roberts *et al.* (2014) for examples of the form of algorithms that would be required).

Secondly, future research may examine the "management paradigm" identified in chapters two and three. This could be performed using alternative sources of data, drawing from a wider pool of disciplines or from the practitioner literature. There is also the potential for empirical studies in this space, potentially investigating the growth of concepts such as analytics amongst practitioners and researchers in the different constituent disciplines in the paradigm. Another direction could include further 'historical' studies of the periods, to include analysis of other artefacts (not just the literature), or interviews with those who experienced different periods. Additionally, follow up studies into the direction of OR research after this study would also be welcome.

In respect to the job advert analysis of chapter four, future research may wish to examine alternative sources of data, potentially from different job sites, using different search keywords, and, in particular, at different times or in a longitudinal fashion. Adaptations to the latent Dirichlet allocation such as the correlated topic model (Blei and Lafferty, 2007) and the topics over time model (Wang and McCallum, 2006) could be utilised to this end. Another potential worthwhile direction, would be replicating, or adapting, the methods here to analyse job adverts in other fields and disciplines.

The analysis of online course materials (chapter five), also suggests opportunities for further investigation. Most obviously, researchers may seek to expand the study internationally, and again extend this from a cross-sectional study to a longitudinal one. Additionally, alternative algorithmic approaches could be employed, potentially treating the problem as continuous rather than discrete (in other words, seeking to identify degrees of similarity rather than specific class prediction). Again, another opportunity would be towards using these methods, or an adaptation thereof, to analyse other disciplines and movements.

For the final instrument, job interviews, there are other relevant directions. Further research would be highly recommended into the requirements of employers and educators more closely aligned to the *data science* end of the analytics spectrum. In a similar vein, and again appropriate to this research but something that the wider research community may wish to address, all interviews were with participants in the UK and Ireland, and not necessarily representative of an international sample.

Overall, similar research would also be recommended in additional areas. Firstly, similar research could seek to analyse the position elsewhere in the world, with the US, as a very large market and also one where courses are taught in English, an obvious direction. Secondly, the work could be extended to evaluate the growing number of undergraduate, or potentially even non-academic, courses in analytics and data science.

In general terms, there are two main cornerstones of what this research has sought to achieve, one methodological and the other thematic. Firstly, methodologically this thesis has consciously sought to apply some of the more recent and novel analytical techniques (topic modelling and ensemble learners for example) in pursuit of meeting its objectives. Given the advantages offered to machine learning in this data-rich age, and the many innovations in the field, we would encourage more researchers in the OR tradition to continue to explore and deploy such methods in their research. Secondly, the topic addressed in this research is one we consider to be high relevance and importance, not just to the OR community, but for both university education and the broader labour market. Particularly considering the pace with which this area is developing and evolving, follow-up research along these lines of this thesis would be strongly encouraged to keep pace with changing requirements of employers, and academic provisions and challenges.

9.4 Concluding Remarks

To return to the argument of chapters two and three, analytics (or indeed data science), are characterised here more as the evolution and latest incarnation of a larger, century-old approach to organisational management rather than the outright paradigmatic shift some of the more enthusiastic proponents of the movement may suggest. However, that does not reduce its impact, nor the scale of the challenge and opportunity it may present. These challenges will be felt in many quarters, but not least for those tasked with recruiting analytics professionals, especially considering the long-term skills shortages associated with scientific and quantitative disciplines.

By implication, this same challenge is also extended to the educators and course designers charged with preparing the next generation of candidates who can fill these roles.

Most tellingly perhaps, now more than ever it seems organisations will require their analytics functions, to some varying extent, to cross all of the key areas of analytics 'stack', from data management and underlying technologies, to the modelling and quantitative approaches, and finally integrating with the systems and people that execute and support organisational decision making. In other words, graduates are, and will increasingly be, called upon to show some degree of mastery of a wide variety of skills and competencies, both "hard" and "soft".

For the OR discipline, it too faces both challenge and opportunity. In respect to the latter, the attention attracted by terms such as "analytics", "data science" and "big data" gives OR a near unprecedented opportunity to reach a wider audience, in particular expanding into new industries and new business functions. However, what this analysis has made clear, is that OR is not the "only show in town", and analytics is clearly the superset of a variety of fields, of which OR is but one. In particular, if computer science and technology schools continue, or increase supply of degrees of the type we describe as *data science* courses in this research, degree courses that may have no mention of OR nor any of its methods, inevitably there is some threat that these become the main source of graduates for the analytics professions. In such a scenario, student numbers for OR and *business analytics* courses would likely fall, as would demand for staff in these areas.

Overall, based on the results of this research, there is enough evidence that points to the value that OR's teachings, curricula and traditions have for employers and the students it produces, and that it is, in most regards, 'fit-for-purpose'. However, that is not to say that OR, and the teaching of *business analytics* courses need not seek to adapt and change to the requirements of analytics.

The origins of OR, most recognisable in the stories of Blackett, Tizard and Bletchley Park, were for the application of a more scientific and logic-based approach to the problems of the day (obviously much of which was the military effort of World War Two). Although OR has come to be associated with specific methods, most obviously optimisation and simulation, OR as it was applied here was essentially technique-agnostic, with the overriding characteristic a focus on an application of scientific methods, to relevant, real world problems. If the recommendations of these research were to be summarised in one single concept, it would be that this same idea be applied to the discipline itself. In other words, the questions OR would need to ask itself, are towards ensuring its problems are the relevant 'problems of the day', and that the tools and techniques employed are the most appropriate available. In doing so, not only can OR remain 'fit-for-purpose', but also maximise the opportunities the interest in analytics may bring, for both the graduates of OR-influenced degrees, and the discipline as a whole.

REFERENCES

Abbott A (1988). *The System of Professions: An essay on the division of expert labor*. University of Chicago Press: Chicago, IL.

Abrahams AS, Jiao J, Fan W, Wang GA and Zhang Z (2013). What's Buzzing in the Blizzard of Buzz? Automotive component isolation in social media postings. *Decision Support Systems* **55:** 871-882.

Ackerman F and Eden C (2010). Strategic Options Development and Analysis. In: Reynolds M and Holwell S (eds.), *Systems Approaches to Managing Change: A practical guide*. Springer: London, pp. 135-190.

Ackoff RL (1956). The Development of Operations Research as a Science. *Operations Research* **4:** 265-295.

Agarwal, R and Dhar V (2014). Editorial—Big Data, Data Science, and Analytics: The opportunity and challenge for IS research. *Information Systems Research* **25:** 443-448.

Agrawal R, Johnson C, Kiernan J and Leymann F (2006). Taming Compliance with Sarbanes-Oxley Internal Controls using Database Technology. *Proceedings of the 22nd International Conference on Data Engineering* (ICDE '06). IEEE, pp. 92-95.

Agosta L (2004). Data Mining is Dead – Long Live Predictive Analytics. *DM Review* **14:** 37-38.

Aizerman MA, Braverman EM and Rozonoer LI (1964). Theoretical Foundations of the Potential Function Method in Pattern Recognition Learning. *Automation and Remote Control* **25:** 821–837.

Anderson C (2008). *The End of Theory: The data deluge makes the scientific method obsolete*. Wired Magazine, 16th July 2008. http://www.wired.com/science/discoveries/magazine/16-07/pb_theory (accessed August 2014).

Andrienko G, Andrienko N, Demsar U, Dransch D, Dykes J, Fabrikant SI, Jern M, Kraak MJ, Schumann H and Tominski C (2010). Space, Time and Visual Analytics. *International Journal of Geographical Information Science* **24:** 1577-1600.

Anthony RN (1965). *Planning and Control Systems: A framework for analysis*. Harvard University Graduate School of Business: Cambridge, MA.

Aral S, Brynjolfsson E and Wu L (2012). Three-way Complementarities: Performance pay, human resource analytics, and information technology. *Management Science* **58:** 913-931.

Armstrong JS, Green KC and Graefe A (2015). Golden Rule of Forecasting: Be conservative. *Journal of Business Research* **68:** 1717-1731.

Asif Z and Mandviwalla M (2005). Integrating the Supply Chain with RFID: A technical and business analysis. *Communications of the Association for Information Systems*, **15:** 393-426.

Aspray W (1990). John von Neumann and the Origins of Modern Computing. MIT Press: Cambridge, MA.

Baier M, Carballo JE, Chang AJ, Yang L, Mojsilovic A, Richard MJ, Singh M, Squillante MS and Varshney KR (2012). Sales-force Performance Analytics and Optimization. *IBM Journal of Research and Development* **56:** 1-10.

Ballings M and Van den Poel D (2012). Customer Event History for Churn Prediction: How long is long enough? *Expert Systems with Applications* **39:** 13517-13522.

Bartlett R (2013). A Practitioner's Guide to Business Analytics: Using data analysis tools to improve your organization's decision making and strategy. McGraw-Hill: New York.

Barton D and Court D (2012). Making Advanced Analytics Work for You. *Harvard Business Review* **90:** 78-83.

Basker E (2012). Raising the Barcode Scanner: Technology and productivity in the retail sector. *American Economic Journal: Applied Economics* **4:** 1-27.

Basu A (2013). Five Pillars of Prescriptive Analytics Success. *Analytics Magazine*, March/April 2013, pp. 8-12.

Beatham S, Anumba C, Thorpe T and Hedges I (2004). KPIs: A critical appraisal of their use in construction. *Benchmarking: An International Journal* **11:** 93-117.

Bell PC and von Lanzenauer CH (2000). Teaching Objectives: The value of using cases in teaching operational research. *Journal of the Operational Research Society*, **51:** 1367-1377.

Belton V and Elder M (1994). Decision Support Systems - Learning from Visual Interactive Modelling. *Decision Support Systems* **12:** 355-364.

Belton V and Scott JL (1998). Independent Learning and Operational Research in the Classroom. *Journal of the Operational Research Society,* **49:** 899-910.

Better M, Glover F and Laguna M (2007). Advances in Analytics: Integrating dynamic data mining with simulation optimization. *IBM Journal of Research and Development* **51:** 477-487.

Bijmolt TH, Leeflang PS, Block F, Eisenbeiss M, Hardie BG, Lemmens A and Saffert P (2010). Analytics for Customer Engagement. *Journal of Service Research* **13:** 341-356.

Birch W (1986). Towards a Model for Problem-based Learning. *Studies in Higher Education* **11:** 73-82.

Birge JR (2006). *Where do we want OR to go in Education and Vice Versa?* http://faculty.chicagobooth.edu/john.birge/research/birge_informs_pgh_or_education_2006nov.pdf (accessed May 2016).

Blackett G (2010). *Analytics Network - O.R. & Analytics.* http://www.theorsociety.com/Pages/SpecialInterest/AnalyticsNetwork_analytics.aspx (accessed April 2014).

Blei DM, Ng AY and Jordan MI (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research* **3:** 993–1022.

Blei DM and Lafferty JD (2007). A Correlated Topic Model of Science. *The Annals of Applied Statistics* **1:** 17-35.

Blight D, Davis D and Olsen A (2000). The Globalization of Higher Education. In: Scott P (ed.), *Higher Education Re-formed*. Routledge: Abingdon and New York, pp.95-113.

Bonomi F, Milito R, Zhu J and Addepalli S (2012). Fog Computing and its Role in the Internet of Things. In: *Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing*. New York: ACM, pp. 13-16.

Boyd D and Crawford (2012). Critical Questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society* **15:** 662-679.

Box GEP and Draper N R (1987). *Empirical Model Building and Response Surfaces*. John Wiley & Sons: New York.

Box GEP and Jenkins GM (1968). Some Recent Advances in Forecasting and Control. *Journal of the Royal Statistical Society*, Series C, **17:** 91-109.

Breiman I (1996). Bagging Predictors. *Machine Learning* **24:** 123–140.

Brown RG (1957). Exponential Smoothing for Predicting Demand. *Operations Research*, **5:** 145-145.

Bruzzone AG and Giribone P (1998). Decision-Support Systems and Simulation for Logistics: Moving forward for a distributed, real-time, interactive simulation environment. *Proceedings of the 31st Annual Simulation Symposium*, 17-24.

Buch N, Velastin SA and Orwell J (2011). A Review of Computer Vision Techniques for the Analysis of Urban Traffic. *IEEE Transactions on Intelligent Transportation Systems* **12:** 920-939.

Buchanan JT, Henig EJ and Henig MI (1998). Objectivity and Subjectivity in the Decision Making Process. *Annals of Operations Research* **80:** 333-345.

Burton I (1963). The Quantitative Revolution and Theoretical Geography. *Canadian Geographer* **7:** 151-162.

Cameron R (2011). Mixed Methods Research: The five Ps framework. *The Electronic Journal of Business Research Methods* **9:** 96-108.

Card SK, Moran TP and Newell A (1983). *The Psychology of Human-Computer Interaction*. Erlbaum: Hilsdale, NJ.

Carneiro HA and Mylonakis E (2009). Google Trends: A web-based tool for real-time surveillance of disease outbreaks. *Clinical Infectious Diseases* **49:** 1557-1564.

Carnell LJ (2008). The Effect of a Student-designed Data Collection Project on Attitudes toward Statistics. *Journal of Statistics Education* **16:** 1-15.

Carter MP (1987). Preliminary Findings of a Survey of OR Society Membership. *The Journal of the Operational Research Society,* **38:** 3-16.

Castillo C, Mendoza M and Poblete B (2011). Information Credibility on Twitter. In: *Proceedings of the 20th International Conference on World Wide Web*. New York: ACM, pp. 675-684.

Cattell R (2010). Scalable SQL and NoSQL Data Stores. *ACM SIGMOD Record* **39:** 12-27.

Ceruzzi PE (1999). *A History of Modern Computing*. The MIT Press: Cambridge, MA and London.

Chao CA and Shih SC (2005). Organizational and End-user Information Systems Job Market: An analysis of job types and skill requirements. *Information Technology, Learning, and Performance Journal* **23:** 1–15.

Chaffey D and Patron M (2012). From Web Analytics to Digital Marketing Optimization: Increasing the commercial value of digital analytics. *Journal of Direct, Data and Digital Marketing Practice* **14:** 30-45.

Champkin J (2014). *The Timeline of Statistics*. https://www.statslife.org.uk/history-of-stats-science/1190-the-timeline-of-statistics (accessed March 2018).

Chang J, Gerrish S, Wang C, Boyd-Graber JL and Blei DM (2009). Reading Tea Leaves: How humans interpret topic models. *Advances in Neural Information Processing Systems 2009* (pp. 288-296).

Charnes A, Cooper WW and Rhodes E (1978). Measuring the Efficiency of Decision Making Units. *European Journal of Operational Research,* **2:** 429-444.

Checkland PB (1981). *Systems Thinking, Systems Practice*. Wiley: Chichester, UK.

Chen H, Chiang RHL and Storey VC (2012). Business Intelligence and Analytics: From Big Data to big impact. *MIS Quarterly* **36:** 1165-1188.

Chen J, Huang H, Tian S and Qu Y (2009). Feature Selection for Text Classification with Naïve Bayes. *Expert Systems with Applications*, **36:** 5432-5435.

Cheung KL and Fu AWC (1998). Enhanced Nearest Neighbour Search on the R-tree. *ACM SIGMOD Record*, **27:** 16-21.

Chiang RHL, Goes P and Stohr EA (2012). Business Intelligence and Analytics Education, and Program Development: A unique opportunity for the Information Systems discipline. *ACM Transactions on Management Information Systems* **3**: 12-24.

Cho CH and Patten DM (2007). The Role of Environmental Disclosures as Tools of Legitimacy: A research note. *Accounting, Organizations and Society* **32:** 639-647.

Christensen CM, Cook S and Hall T (2005). Marketing Malpractice: The cause and the cure. *Harvard Business Review* **83:** 74-83.

Chu H (2006). Curricula of LIS programs in the USA: A content analysis. In: Khoo C, Singh D and Chaudhry AS (eds.), *Proceedings of the Asia-Pacific Conference on Library & Information Education & Practice 2006*, 3-6 April 2006. Singapore: Nanyang Technological University, pp. 328-337.

Churchman C West (1967). Wicked Problems. *Management Science* **14:** B141-B142.

Churchman C West (1970). Operations Research as a Profession. *Management Science* **17:** B-37-B53.

Cochran JJ (2009). Pedagogy in Operations Research: Where has the discipline been, where is it now, and where should it go? *ORiON* **25:** 161:184.

Cohen J, Dolan B, Dunlap M, Hellerstein JM and Welton C (2009). MAD Skills: New analysis practices for Big Data. *Proceedings of VLDB Endowment* **2:** 1481-1492.

Cook J (2013). *Shares of Tableau Still Surging After Strong IPO, Now Valued Near $2B.* GeekWire http://www.geekwire.com/2013/shares-tableau-surging-strong-ipo-valued-2b/ (accessed March 2013).

Corbett CJ and Van Wassenhove LN (1993). The Natural Drift: What happened to operations research? *Operations Research* **41:** 625-640.

Corne D, Dhaenens C and Jourdan L (2002). Synergies between Operations Research and Data Mining: The emerging use of multi-objective approaches. *European Journal of Operational Research* **221:** 469–479.

Cortes C and Vapnik V (1995). Support-vector Networks. *Machine Learning* **20:** 273-297.

D'Haen J, Van den Poel D and Thorleuchter D (2013). Predicting Customer Profitability During Acquisition: Finding the optimal combination of data source and data mining technique. *Expert Systems with Applications* **40:** 2007-2012.

Dando MR and Bennett PG (1981). A Kuhnian Crisis in Management Science? *Journal of the Operational Research Society* **32:** 91-103.

Danson F, Amjad A, Davenport T, Gold S, Guszcza J, Katyal V, Lucker J, Mueller B and Raphael J (2016). *Analytics trends 2016: The next evolution.* http://www2.deloitte.com/content/dam/html/us/analytics-trends/2016-analytics-trends/pdf/analytics-trends.pdf (accessed February 2016).

Date CJ (2000). Database Relational Model: A retrospective review and analysis. Addison Wesley, Reading, MA.

Davenport TH, Barth P and Bean R (2012). How 'Big Data' is Different. *MIT Sloan Management Review* **54:** 22-24.

Davenport TH and Harris JG (2007). *Competing on Analytics: The new science of winning*. Harvard Business School: Boston, MA.

Davenport TH and Patil DJ (2012). Data Scientist: The sexiest job of the 21st Century. *Harvard Business Review* **90:**70–76.

Davidow W and Malone M (1992). The Virtual Corporation - Structuring and Revitalising the Corporation for the 21st Century. HarperCollins: New York.

Davis G (1982). Management Information Systems: A fifteen-year perspective. *ACM SIGMS Database* **13:** 10-11.

Davis WJ and Jones AT (1988). A Real-Time Production Scheduler for a Stochastic Manufacturing Environment. *International Journal of Computer Integrated Manufacturing* **1:** 101-112.

Davy M and Luz S (2007). Dimensionality Reduction for Active Learning with Nearest Neighbour Classifier in Text Categorisation Problems. In: *Sixth International Conference on Machine Learning and Applications, 2007*. ICMLA 2007 (pp. 292-297). IEEE.

Dean J and Ghemawat S (2004). MapReduce: Simplified data processing on large clusters. In: *Proceedings of the 6th Conference on Symposium on Operating Systems Design and Implementation*, San Francisco.

Den Hartog DN, Caley A and Dewe P (2007). Recruiting Leaders: An analysis of leadership advertisements. *Human Resources Management Journal* **17:** 58–75.

Devlin BA and Murphy PT (1988). An Architecture for a Business and Information System. *IBM Systems Journal* **27:** 60-80.

Dhar V (2013). Data Science and Prediction. *Communications of the ACM* **12:** 64-73.

Di Domenica N, Mitra G, Valente P and Birbilis G (2007). Stochastic Programming and Scenario Generation within a Simulation Framework: An information systems perspective. *Decision Support Systems* **42:** 2197-2218.

Diamond RM (2008). *Designing and Assessing Courses and Curricula: A practical guide* (3rd edition – Kindle version). Jossey Bass: San Francisco.

Diehl M, Bock HG, Schloder JP, Findeisen R, Nagy Z, Allgower F (2002). Real-Time Optimization and Nonlinear Model Predictive Control of Processes Governed by Differential-Algebraic Equations. *Journal of Process Control* **12:** 577–585.

Domingos P (2012). A Few Useful Things to Know about Machine Learning. *Communications of the ACM* **55:** 78-87.

Driscoll K (2012). From Punched Cars to 'Big Data': A social history of database populism. *communication +1* **4:** 1-33. http://scholarworks.umass.edu/cpo/vol1/iss1/4 (accessed January 2013).

Dumais ST (2004). Latent Semantic Analysis. *Annual Review of Information Science and Technology*, **38:** 188-230.

Durbin M and Hoffman K (2008). The Dance of the Thirty-Ton Trucks: Dispatching and scheduling in a dynamic environment. *Operations Research* **56:** 3-19.

Dyson G, Kelly K, Brand S, Hills DW, Carroll S, Lanier J, Traub J, Horgan J, Sterling B, Rushkoff D, Morton O, Everett D, Origgi G, Smolin L and Garreau J (2008). *On Chris Anderson's the End of Theory*. Edge. http://www.edge.org/discourse/the_end_of_theory.html (accessed May 2013).

e-Skills and SAS (2013). *Big Data Analytics: Adoption and employment trends, 2012-2017.* http://www.sas.com/offices/europe/uk/downloads/bigdata/eskills/eskills.pdf (accessed July 2015).

Eckerson W (2011). *What is Analytics?* BeyeNETWORK. http://www.b-eye-network.com/blogs/eckerson/archives/2011/07/what_is_analyti.php (accessed July 2015).

Eirinaki M and Vazirgiannis M (2003). Web Mining for Web Personalization. *ACM Transactions on Internet Technology* **3:** 1-27.

Efron B (1979). Computers and the Theory of Statistics: Thinking the unthinkable. *SIAM review*, **21:** 460-480.

Elbashir MZ, Collier PA and Sutton SG (2011). The Role of Organizational Absorptive Capacity in Strategic use of Business Intelligence to Support Integrated Management Control Systems. *The Accounting Review* **86:** 155-184.

Elliot T (2011). *Business Analytics vs. Business Intelligence?* http://timoelliott.com/blog/2011/03/business-analytics-vs-business-intelligence.html (accessed July 2015).

Esch JW (1969). *RASCEL: A programmable analog computer based on a regular array of stochastic computing element logic* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (Accession Order No. 7000843).

Evans JR (2012). Business Analytics: The next frontier for decision sciences. *Decision Line* 43: 4-6.

Fan R-E, Chang K-W, Hsieh C-J, Wang X-R and Lin C-J (2008). LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research* **9:** 1871-1874.

Fayyad U, Piatetsky-Shapiro G and Smyth P (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, Fall 1996, pp. 37-54.

Feinleib D (2012). *The Big Data Landscape*. http://www.forbes.com/sites/davefeinleib/2012/06/19/the-big-data-landscape/ (accessed March 2013).

Ferguson R (2012). Learning Analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning* **4:** 304-317.

Ferguson RL and Jones CH (1969). A Computer Aided Decision System. *Management Science* **15:** B550-B561.

Few S (2006). Information Dashboard Design: The effective communication of data. O'Reilly: Beijing and Cambridge, MA.

FICO (2013). *Large-Scale Marketing Campaign Optimization Using FICO® Xpress Optimization Suite to Support Multichannel Campaign Management.* http://www.fico.com/en/FIResourcesLibrary/Large_Scale_Marketing_Campaign_Optimization_2946WP_EN.pdf (accessed November 2013).

Fildes R and Ranyard JC (1997). Success and Survival of Operational Research Groups: A review. *The Journal of the Operational Research Society*, **48:** 336-360.

Fildes R, Ranyard JC and Crymble WR (1999). The Management of OR Groups: Results of a survey .*The Journal of the Operational Research Society*, **50:** 563-580.

Flowers TH (1983). The Design of Colossus. *Annals of the History of Computing* **5:** 38-48.

Frank E and Bouckaert RR (2006). Naive Bayes for Text Classification with Unbalanced Classes. In: *European Conference on Principles of Data Mining and Knowledge Discovery* (pp. 503-510). Springer: Berlin, Heidelberg.

Freakonomics (2008). *Hal Varian Answers Your Questions*. http://freakonomics.com/2008/02/25/hal-varian-answers-your-questions/ (accessed March 2018).

Friendly M (1995). Milestones in the History of Thematic Cartography, Statistical Graphics, and Data Visualization. *13th International Conference on Database and Expert Systems Applications*.

Fu G, Kapelan Z, Kasprzyk JR and Reed P (2012). Optimal Design of Water Distribution Systems using Many-objective Visual Analytics. *Journal of Water Resources Planning and Management* **139:** 624-633.

Genkin A, Lewis DD and Madigan D (2007). Large-scale Bayesian Logistic Regression for Text Categorization. *Technometrics*, **49:** 291-304.

Germann F, Lilien GL and Rangaswamy A (2013). Performance Implications of Deploying Marketing Analytics. *International Journal of Research in Marketing* **30:** 114-128.

Giaglis GM, Minis I, Tatarakis A and Zeimpekis V (2004). Minimizing Logistics Risk through Real-Time Vehicle Routing and Mobile Technologies: Research to date and future trends. *International Journal of Physical Distribution & Logistics Management* **34:** 749-764.

Giloi WK (1998). Konrad Zuse: Reflections on the 80th Birthday of the German computing pioneer. *ACM SIGNUM* **33:** 11-16.

Giuliani M., Galelli S and Soncini-Sessa R (2014). A Dimensionality Reduction Approach for Many-objective Markov Decision Processes: Application to a water reservoir operation problem. *Environmental Modelling & Software* **57:** 101-114.

Gould R (2010). Statistics and the Modern Student. *International Statistical Review* **78:** 297-315.

Granville V (2013). *The End of Theory: The data deluge makes the scientific method obsolete | Wired*. Analytic Bridge. http://www.analyticbridge.com/profiles/blogs/the-end-of-theory-the-data-deluge-makes-the-scientific-method-obs (accessed May 2013).

Gray PJ, Diamond RM and Adam BE (1996). A National Study on the Relative Importance of Research and Undergraduate Teaching at Colleges and Universities. Center for Instructional Development, Syracuse University.

Green WA (1995). Periodizing World History. *History and Theory* **34:** 99-111.

Griffiths TL, Jordan MI, Tenenbaum JB and Blei DM (2004). Hierarchical Topic models and the Nested Chinese Restaurant Process. *Advances in Neural Information Processing Systems*, volume 17. MIT Press: Cambridge, MA (pp. 537–544).

Grossman TA (2001). Causes of the Decline of the Business School Management Science Course. *INFORMS Transactions on Education* **1:** 51-61.

Grossman TA, Mehrotra V and Sidaoui M (2016) A Student-Centered Approach to the Business School Management Science Course. *INFORMS Transactions on Education* **16:** 42-53.

Gürer D (2002). Women in Computing. *ACM SIGCSE Bulletin - Women and Computing* **34:** 116-120.

Hämälläinen RP, Luoma J and Saarinen E (2013). On the Importance of Behavioral Operational Research: The Case of Understanding and Communicating about Dynamic Systems. European Journal of Operational Research 228, 623–624.

Hammerbacher J (2009). Information Platforms and the Rise of the Data Scientist. In: Segaran T and Hammerbacher J (eds.), *Beautiful Data: The stories behind elegant data solutions*. O'Reilly: Sebastopol, CA, pp. 73-84.

Han J, Kamber M and Pei J (2011). *Data Mining: Concepts and techniques*. Waltham, MA: Elsevier.

Hardin J, Hoerl R, Horton NJ, Nolan D, Baumer B, Hall-Holt O, Murrell P, Peng R, Roback P, Temple Lang D and Ward MD (2015). Data Science in Statistics Curricula: Preparing students to "think with data". *The American Statistician* **69:** 343-353.

Hauser W J (2007). Marketing Analytics: The evolution of marketing research in the twenty-first century. *Direct Marketing: An International Journal* **1:** 38-54.

He J (2009). Advances in Data Mining: History and future. Proceedings of the 3rd International Symposium on Intelligent Information Technology Application, pp. 634-636.

Helbing D and Balietti S (2011). *From Social Data Mining to Forecasting Socio-Economic Crises*. https://arxiv.org/pdf/1012.0178v5.pdf (accessed March 2018).

Henderson KM and Evans JR (2000). Successful Implementation of Six Sigma: Benchmarking General Electric Company. *Benchmarking: An International Journal* **7:** 260–282.

HM Government (2013). *Seizing the Data Opportunity: A strategy for UK data capability*. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/254136/bis-13-1250-strategy-for-uk-data-capability-v4.pdf (accessed June 2015).

Hochreiter S and Schmidhuber J (1997). Long Short-term Memory. *Neural Computation*, **9:** 1735-1780.

Hoffman M, Bach FR and Blei DM (2010). Online Learning for Latent Dirichlet allocation. *Advances in Neural Information Processing Systems 2010* (pp. 856-864).

Hofmann T (1999). Probabilistic Latent Semantic Indexing. In: *Proceedings of the Twenty-Second Annual International SIGIR Conference on Research and Development in Information Retrieval (SIGIR-99).*

Hookway C (2016). Pragmatism. In: Zalta EN (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2016 Edition). http://plato.stanford.edu/archives/sum2016/entries/pragmatism (accessed July 2016).

Hurrion RD (1976). The Design, Use and Required Facilities of an Interactive Visual Computer Simulation Language to Explore Production Planning Problems. PhD thesis, University of London.

IBM (2013). *Bringing Big Data to the Enterprise*. http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html (accessed May, 2016).

Iggers GG (2005). Historiography in the Twentieth Century: From scientific objectivity to the postmodern challenge. Wesleyan University Press: Middletown, CT.

INFORMS (2013). *What is Operations Research?* https://www.informs.org/About-INFORMS/What-is-Operations-Research (accessed July 2015).

INFORMS (2014). *Certified Analytics Professional: Candidate handbook*, 3rd Edition. https://www.informs.org/content/download/266243/2511357/file/CAP_handbookV3.pdf (accessed August 2014).

Isenberg P, Fisher D, Paul SA, Morris MR, Inkpen K and Czerwinski M (2012). Co-located Collaborative Visual Analytics around a Tabletop Display. *IEEE Transactions on Visualization and Computer Graphics* **18:** 689-702.

Jacobs FR and Weston FC (2007). Enterprise Resource Planning (ERP): A brief history. *Journal of Operations Management*, **25:** 357-363.

Janiesch C, Matzner M and Müller O (2012). Beyond Process Monitoring: A proof-of-concept of event-driven business activity management. *Business Process Management Journal* **18:** 625-643.

Joachims T (1998). Text Categorization with Support Vector Machines: Learning with many relevant features. In *European Conference on Machine Learning*. Berlin: Springer (pp. 137-142).

Johnson C (2012). *Advancing Analytics to Predict Specific Needs*. A Smarter Planet (IBM). http://asmarterplanet.com/blog/2012/08/advancing_analytics.html (accessed June 2015).

Johnson RB and Onwuegbuzie AJ (2004). Mixed Methods Research: A research paradigm whose time has come. *Educational Researcher* **33:** 14-26.

Johnson RB, Onwuegbuzie AJ and Turner LA (2007). Towards a Definition of Mixed Methods Research. *Journal of Mixed Methods Research* **1:** 112-133.

Joshi M, Das D, Gimpel K and Smith NA (2010). Movie Reviews and Revenues: An experiment in text regression. In: *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 293-296).

Kaplan EH (2008). Adventures in Policy Modeling! Operations research in the community and beyond. *Omega* **36:** 1-9.

Kaplan EL and Meier P (1958). Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, **53:** 457-481.

Kaplan RS and Norton DP (1992). The Balanced Scorecard – Measures that Drive Performance. *Harvard Business Review* **71:** 71-79.

Kapoor B (2010). Business Intelligence and its use for Human Resource Management. *The Journal of Human Resource and Adult Learning* **6:** 21-30.

Kardes FR, Cronley ML and Cline TW (2010). *Consumer Behavior*. South-Western Cengage Learning: Mason, OH.

Kasprzyk JR, Nataraj S, Reed PM and Lempert RJ (2013). Many Objective Robust Decision Making for Complex Environmental Systems Undergoing Change. *Environmental Modelling & Software* **42:** 55-71.

Kasprzyk JR., Reed PM, Characklis GW and Kirsch BR (2012). Many-objective de Novo Water Supply Portfolio Planning under Deep Uncertainty. *Environmental Modelling & Software* **34:** 87-104.

Khandani AE, Kim AJ and Lo AW (2010). Consumer Credit-risk Models via Machine-learning Algorithms. *Journal of Banking & Finance* **34:** 2767-2787.

King N (2004). Using Templates in the Thematic Analysis of Texts. In: Cassell C and Symon G (eds.), *Essential Guide to Qualitative Methods in Organizational Research*, pp. 256-270.

Kirby MW (2003). *Operational Research in War and Peace*. Imperial College Press: London; and The Operational Research Society: Birmingham, UK.

Kitchin R (2014). The Real-time City? Big data and smart urbanism. *GeoJournal* **79:** 1-14.

Kivinen J, Warmuth MK and Auer P (1997). The Perceptron Algorithm versus Winnow: Linear versus logarithmic mistake bounds when few input variables are relevant. *Artificial Intelligence*, **97:** 325-343.

Klatt T, Schlaefke M and Moeller K (2011). Integrating Business Analytics into Strategic Planning for Better Performance. *Journal of Business Strategy* **32:** 30-39.

Knuth DE (1968). *The Art of Computer Programming*, 1st edition. Addison-Wesley: Reading, MA.

Kohavi R, Rothleder NJ and Simoudis E (2002). Emerging Trends in Business Analytics. *Communications of the ACM* **45:** 45-48.

Köksalan M, Wallenius J and Zionts S (2011). *Multiple Criteria Decision Making: From early history to the 21st Century*. World Scientific. Singapore.

Kloss C, Goniva C, Hager A, Amberger S and Pirker S (2012). Models, Algorithms and Validation for Opensource DEM and CFD–DEM. *Progress in Computational Fluid Dynamics, an International Journal* **12:** 140-152.

Krämer A, Green J, Pollard J and Tugendreich S (2014). Causal Analysis Approaches in Ingenuity Pathway Analysis. *Bioinformatics* **30:** 523-530.

Krause J, Perer A and Bertini E (2014). INFUSE: Interactive feature selection for predictive modeling of high dimensional data. *IEEE Transactions on Visualization and Computer Graphics* `1614-1623.

Kuhn T (1962). *The Structure of Scientific Revolutions*. University of Chicago Press: Chicago.

Laney D (2001). *3D Data Management: Controlling data volume, velocity, and variety*. http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf (accessed March 2014).

Laney D and Kart L (2012). *Emerging Role of the Data Scientist and the Art of Data Science*. Gartner. https://www.gartner.com/doc/1955615/emerging-role-data-scientist-art (accessed November 2015).

Laporte G (2008). Operational Research: Milestones and highlights of Canadian contributions. *INFOR* **46:** 309-312.

Laursen GHN and Thorlund J (2010). Business Analytics for Managers: Taking business intelligence beyond reporting. Wiley. Hoboken, NJ.

LeCun Y, Bottou L, Orr G and Muller M (1998). Efficient Backprop. In Orr G and Muller K (eds.) *Neural Networks: Tricks of the trade*. Springer: Berlin (pp. 9-48).

Lee M and Mimno D (2017). Low-dimensional Embeddings for Interpretable Anchor-based Topic Inference. *arXiv preprint*, arXiv:1711.06826.

Lesser LM (2007). Critical Values and Transforming Data: Teaching statistics with social justice. *Journal of Statistics Education* **15:** 1-21.

Lewis J (1996). Reaping the Benefits of Modern Usability Evaluation: The Simon story. *1st International Conference on Advances in Applied Ergonomics*, pp. 752-757.

Liberatore M and Luo W (2010). The Analytics Movement: Implications for operations research. *Interfaces* **40**: 313-324.

Liberatore M and Luo W (2011). INFORMS and the Analytics Movement: The view of the membership. *Interfaces* **41:** 578-589.

Liberatore M and Luo W (2013). ASP, The Art and Science of Practice: A comparison of technical and soft skill requirements for analytics and OR professionals. *Interfaces* **43:** 194-197.

Liberatore MJ and Nydick RL (1999) The Teachers' Forum: Breaking the mold – a new approach to teaching the first MBA course in management science. *Interfaces* **29:** 99-116.

Lim EP, Chen H and Chen G (2012). Business Intelligence and Analytics: Research directions. *ACM Transactions on Management Information Systems* **3:** 17-27.

Lior R (2014). *Data Mining with Decision Trees: Theory and applications*. World Scientific: Singapore.

Liu B and Zhang L (2012) A Survey of Opinion Mining and Sentiment Analysis. In: Aggarwal C and Zhai C (eds.) *Mining Text Data*. Springer: Boston, MA.

Liu S, Zhou MX, Pan S, Qian W, Cai W and Lian X (2009). Interactive, Topic-Based Visual Text Summarization and Analysis. In: *Proceedings of the 18th ACM Conference on Information and Knowledge Management.* New York: ACM, pp. 543-552.

Locke RR (1989). *Management and Higher Education since 1940*. Cambridge University Press: Cambridge, UK.

Lonn S, Aguilar SJ and Teasley SD (2015). Investigating Student Motivation in the Context of a Learning Analytics Intervention during a Summer Bridge Program. *Computers in Human Behavior* **47:** 90-97.

Lunt BM, Ekstrom JJ, Gorka S, Hislop G, Kamali R, Lawson E, LeBlanc R, Miller J and Reichgelt H (2008). *Curriculum Guidelines for Undergraduate Degree Programs in Information Technology*. ACM and IEEE-CS. ftp://182.255.2.82/pustaka/kurikulum-acm/IT2008%20Curriculum_2.pdf (accessed June 2014).

Lustig I, Dietrich B, Johnson C and Dziekan C (2010). The Analytics Journey. *Analytics Magazine*, November/December 2010, pp. 11-13.

McAfee A and Brynjolfsson E (2012). Big data. The management revolution. *Harvard Business Review* **90:** 61-67.

Mahadev A and Wurst KR (2015). Developing Concentrations in Big Data Analytics and Software Development at a Small Liberal Arts University. *Journal of Computing Sciences in Colleges* **30:** 92-98.

Malnik M (2012). Figures of Prosleptic Syllogisms in Prior Analytics 2.7. *The Classical Quarterly* **62:** 163-178.

Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C and Hung Byers A (2011). *Big Data: The next frontier for innovation, competition and productivity*. McKinsey Global Institute. http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation (accessed April 2016).

Marcus A (2006). Dashboards in your Future. *Interactions – The Art of Prototyping*. January/February 2006, pp. 48-60.

Marshall MN (1996). Sampling for Qualitative Research. *Family Practice* **13:** 522-525.

Maslow AH (1969). *Psychology of Science*. Regnery: Washington D.C.

Masada T, Kiyasu S and Miyahara S (2008) Comparing LDA with pLSI as a Dimensionality Reduction Method in Document Clustering. In: Tokunaga T and Ortega A (eds.). *Large-Scale Knowledge Resources. Construction and Application. LKR 2008. Lecture Notes in Computer Science*, vol. 4938. Springer: Berlin, Heidelberg.

Mass R (2018). *Alexa Scientists Address Challenges of End-Pointing.* https://developer.amazon.com/blogs/alexa/post/bb1909e9-aa65-4051-9316-eba9dd3bf459/alexa-scientists-address-challenges-of-end-pointing (accessed May 2018).

Mathews D and Clark J (2003). *Successful Students' Conceptions of Mean, Standard Deviation, and the Central Limit Theorem.* https://www.researchgate.net/profile/David_Mathews4/publication/253438034_Successful_Students'_Conceptions_of_Mean_Standard_Deviation_and_The_Central_Limit_Theorem/links/54e23d6b0cf2c3e7d2d335af.pdf (accessed May 2016).

Mathioudakis M and Koudas N (2010). TwitterMonitor: Trend detection over the twitter stream. In: *Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data*. New York: ACM, pp. 1155-1158.

Mayer-Schönberger V and Cukier K (2013). Big Data: A revolution that will transform how we live, work and think. John Murray: London.

McAuliffe JD and Blei DM (2008). Supervised Topic Models. In *Advances in Neural Information Processing Systems* (pp. 121-128).

Metropolis N (1989). The Beginning of the Monte Carlo Method. In: Cooper NG, Eckhardt R and Shera N (eds). *From Cardinals to Chaos: Reflections on the life and legacy of Stanislaw Ulam*. New York: Cambridge University Press, pp. 125–130.

Metters R, Queenan C, Ferguson M, Harrison L, Higbie J, Ward S, Barfield B, Farley T, Kuyumcu HA and Duggasani A (2008). The "Killer Application" of Revenue Management: Harrah's Cherokee Casino & Hotel. *Interfaces* **38:** 161–175.

Mian R, Martin P and Vazquez-Poletti JL (2013). Provisioning Data Analytic Workloads in a Cloud. *Future Generation Computer Systems* **29:** 1452-1458.

Miguéis VL, Van den Poel D, Camanho AS and e Cunha JF (2012). Modeling Partial Customer Churn: On the value of first product-category purchase sequences. *Expert Systems with Applications* **39:** 11250-11256.

MIKE 2.0 (2012). *Big Data Definition.* http://mike2.openmethodology.org/wiki/Big_Data_Definition (accessed November 2014).

Miles MB and Huberman MA (1994). *Qualitative Data Analysis: An expanded sourcebook* (2nd edition). Sage: Thousand Oaks, CA.

Mimno D, Wallach HM, Talley E, Leenders M and McCallum A (2011). Optimizing Semantic Coherence in Topic Models. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (pp. 262-272). Association for Computational Linguistics.

Mithas S, Lee MR, Earley S, Murugesan S and Djavanshir R (2013). Leveraging Big Data and Business Analytics. *IT professional*, **15:** 18-20.

Molluzzo JC, Lawler JP and Plaza OP (2015). A Proposed Concentration Curriculum Design for Big Data Analytics for Information Systems Students. *Information Systems Education Journal* **13:** 45-57.

Moore DS (1997). New Pedagogy and New Content: The case of statistics. *International Statistical Review* **65:** 123-165.

Mortenson MJ, Doherty NF and Robinson S (2015). Operational Research from Taylorism to Terabytes: A research agenda for the analytics age. *European Journal of Operational Research* **241:** 583-595.

Mortenson MJ and Vidgen R (2015). A Computational Literature Review of the Technology Acceptance Model. *International Journal of Information Management* (forthcoming).

Mun J (2010). Modeling Risk: Applying Monte Carlo risk simulation, strategic real options, stochastic forecasting, and portfolio optimization. John Wiley & Sons: Hoboken (NJ).

Mutschler C, Ziekow H and Jerzak Z (2013). The DEBS 2013 Grand Challenge. In: *Proceedings of the 7th ACM International Conference on Distributed Event-based Systems* New York: ACM, pp. 289-294.

Nadin S and Cassell C (2004). Using Data Matrices. In: Cassell C and Symon G (eds.), *Essential Guide to Qualitative Methods in Organizational Research*, pp. 271-287.

Nair H, Misra S, Hornbuckle WJ, Mishra R and Acharya A (2013). *Big Data and Marketing Analytics in Gaming: Combining empirical models and field experimentation*. Stanford University Graduate School of Business Research Paper No. 14-07. http://ssrn.com/abstract=2399676 (accessed April 2016).

Nelson Ford F (1985). Decision Support Systems and Expert Systems: A comparison. *Information & Management* **8:** 21-26.

Newbold P (1975). The Principles of the Box-Jenkins Approach. *Operational Research Quarterly*, **26:** 397-412.

Ng K, Ghoting A, Steinhubl SR, Stewart WF, Malin B and Sun J (2014). PARAMO: A PARAllel Predictive MOdeling Platform for Healthcare Analytic Research using Electronic Health Records. *Journal of Biomedical Informatics* **48:** 160-170.

Niedermann F, Radeschütz S and Mitschang B (2011). Business Process Optimization Using Formalized Optimization Patterns. In: van Sinderen M (ed.), *Business Information Systems: Lecture notes in business information processing* **76:** 123-135. Springer: Berlin.

Niwattanakul S, Singthongchai J, Naenudorn E and Wanapu S (2013). Using of Jaccard coefficient for keywords similarity. *Proceedings of the International Multi-Conference of Engineers and Computer Scientists*.

Nylund AL (1999). *Tracing the BI Family Tree. Knowledge Management*. http://www.escholar.com/documents/dw_family_tree.pdf (accessed January 2013).

Olafsson S, Li X and Wu S (2008). Operations Research and Data Mining. *European Journal of Operational Research* **187:** 1429–1448.

Ormerod R (1999). Blackett the Father of OR. *OR Insight* **12:** 14-20.

Pang B, Lee L and Vaithyanathan S (2002). Thumbs up? Sentiment classification using machine learning techniques. *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (EMNLP), pp. 79–86.

Panian Z (2008). A Break-Through Approach to Real-Time Decisioning in Business Management. In: *Proceedings of the 2nd WSEAS International Conference on Management, Marketing and Finances*, pp. 25-30.

Pease G, Byerly B and Fitz-enz J (2012). Human Capital Analytics: How to harness the potential of your organization's greatest asset. John Wiley & Sons: Hoboken (NJ).

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, and Duchesnay E (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* **12:** 2825 - 2830.

Peterson M, Gröne F, Kammer K and Kirscheneder J (2010). Multi-Channel Customer Management: Delighting consumers, driving efficiency. *Journal of Direct, Data and Digital Marketing Practice* **12:** 10-15.

Phippen A, Sheppard L and Furnell S (2004). A Practical Evaluation of Web Analytics. *Internet Research* **14:** 284-293.

Pidd M (1999). Just Modeling Through: A rough guide to modeling. *Interfaces* **29:** 118-132.

Powell WB, Marar A, Gelfand J and Bowers S (2002). Implementing Real-Time Optimization Models: A case application from the motor carrier industry. *Operations Research* **50:** 571-581.

Power DJ (2007). *A Brief History of Decision Support Systems*. DSS Resources. http://dssresources.com/history/dsshistory.html (accessed December 2012).

Provost F and Fawcett T (2013). Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data* **1:** 51-59.

Pugh E (1984). Ferrite Core Memories that Shaped an Industry. *IEEE Transactions on Magnetics* **20:** 1499-1502.

PyData (2012). *Pandas*. http://pandas.pydata.org/index.html (accessed March 2015).

Rand G and Ranyard J (2013). *Preparing Students for the World of OR Work: Lancaster's experience.* Lancaster University Management School. http://www.research.lancs.ac.uk/portal/en/publications/preparing-students-for-the-world-of-or-work(0e85b47d-cbee-469a-8786-5d63516a0cc6)/export.html (accessed April 2018).

Randell B (1972). On Alan Turing and the Origins of Digital Computers. In: Meltzer B and Michie D (eds.), *Machine Learning* **7:** 3-20.

Ranyard JC, Fildes R and Hu TI (2015). Reassessing the Scope of OR Practice: The influences of problem structuring methods and the analytics movement. *European Journal of Operational Research*, **245:** 1-13.

Redman T and Mathews BP (1997). What do Recruiters Want in a Public Sector Manager? *Public Personnel Management* **26:** 245–256.

Rees M (1980). The Mathematical Science and World War II. *The American Mathematical Monthly* **87:** 607-621.

Rehurek R and Sojka P (2010). Software Framework for Topic Modelling with Large Corpora. *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. ELRA: Valetta, Malta, 22nd May 2010 (pp. 45-50).

Reinsel D, Gantz J and Rydning J (2017). *Data Age 2025: The evolution of data to life-critical.* https://www.seagate.com/www-content/our-story/trends/files/Seagate-WP-DataAge2025-March-2017.pdf (accessed June 2018).

Renders J-M (2004). Kernel Methods for Natural Language Processing. In: *Learning Methods for Text Understanding and Mining*. PASCAL. January 26-29, 2004. Grenoble: Xerox Research Centre Europe.

Roberts ME, Brandon MS, Tingley D, Lucas C, Leder-Luis J, Gadarian SK, Albertson B and Rand DG (2014). Structural Topic Models for Open-Ended Survey Responses. *American Journal of Political Science*, **58:** 1064-1082.

Robins RH (1997). *A Short History of Linguistics* (4th edition). Routledge: Abingdon, UK.

Robinson S, Meadows M, Mingers J, O'Brien FA, Shale EA and Stray S (2003) Teaching OR/MS to MBAs at Warwick Business School: A turnaround story. *Interfaces* **33:** 67-76.

Rockart JF (1979). Chief Executives Define Their Own Data Needs. *Harvard Business Review* **67:** 81-93.

Rosenhead J and Mingers J (2001). Rational Analysis for a Problematic World Revisited: Problem structuring methods for complexity, uncertainty and conflict. Wiley: Chichester, UK.

Royston G (2013). Operational Research for the Real World: Big questions from a small island. *Journal of the Operational Research Society* **64:** 793-804.

Saaty TL (1980). *The Analytic Hierarchy Process*. McGraw-Hill: New York.

Sahay BS and Ranjan J (2008). Real Time Business Intelligence in Supply Chain Analytics. *Information Management & Computer Security* **16:** 28-48.

Sand G and Engell S (2004). Modeling and Solving Real-Time Scheduling Problems by Stochastic Integer Programming. *Computers and Chemical Engineering* **28:** 1087–1103.

Sanders MA (2013). *Data Science - The Process of Capturing, Analyzing and Presenting Business Intelligence with Skill.* http://www.datasciencecentral.com/profiles/blogs/data-scientist-core-skills (accessed April 2016).

SAP (2012). *Empower Your People to Achieve Remarkable Results with Business Analytics Solutions.* http://www.sap.com/solutions/analytics/why-sap-analytics.epx (accessed April 2016).

Sau TK., Rogach AL, Jäckel F, Klar TA and Feldmann J (2010). Properties and Applications of Colloidal Nonspherical Noble Metal Nanoparticles. *Advanced Materials* **22:** 1805-1825.

Schlegel K and Hare J (2017). *Hype Cycle for Analytics and Business Intelligence, 2017*. Gartner. https://www.gartner.com/doc/3772080 (accessed March 2018).

Schmid H (1994). Probabilistic Part-of-speech Tagging using Decision Trees. In: New Methods in Language Processing, Manchester, UK, 1994 (p. 154).

Scott Morton MS (1969). *Computer-Driven Visual Display Devices - Their Impact on the Management Decision-Making Process* (Doctoral thesis). http://dspace.mit.edu/bitstream/handle/1721.1/48089/impactofinteract00scot.pdf?sequence=1 (accessed March 2013).

Seguin R, Potvin J-Y, Gendreau M, Crainic TG, Marcotte P (1997). Real-Time Decision Problems: An operational research perspective. *Journal of the Operational Research Society* **48:** 162-174.

Senn PR (1966). What is "Behavioral Science" – Notes Toward a History. *Journal of the History of the Behavioral Sciences* **2:** 107-122.

Shackelford R, McGettrick A, Sloan R, Topi H, Davies G, Kamali R, Cross J, Impagliazzo J, LeBlanc R and Lunt B (2006). Computing Curricula 2005: The overview report. *ACM SIGCSE Bulletin* **38:** 456-457.

Shah S, Horne A and Capellá J (2013). Good Data Won't Guarantee Good Decisions. *Harvard Business Review* **90:** 23-25.

Shahnaz F, Berry MW, Pauca VP and Plemmons RJ (2006). Document Clustering using Nonnegative Matrix Factorization. *Information Processing & Management*, **42:** 373-386.

Shannon RE and Biles WE (1970). The Utility of Certain Curriculum Topics to Operations Research Practitioners. *Operations Research* **18:** 741-745.

Shim J, Warkentin M, Courtney J, Power D, Sharda R and Carlsson C (2002). Past, Present, and Future of Decision Support Technology. *Decision Support Systems* **33:** 111-126.

Shin SJ, Woo J and Rachuri S (2014). Predictive Analytics Model for Power Consumption in Manufacturing *Procedia CIRP* **15:** 153-158.

Shmueli G and Koppius OR (2011). Predictive Analytics in Information Systems Research. *MIS Quarterly* **35:** 553-572.

Simon HA (1947). *Administrative Behavior*. Macmillan: New York.

Simon HA (1965). *The New Science of Management Decision*. Prentice-Hall: Englewood Cliffs, NJ.

Smil V (2005). Creating the Twentieth Century: Technical innovations of 1867-1914 and their lasting impact. Oxford University Press: Oxford and New York.

Smith J and Meaney M (2016). *Lifelong Learning in 2040: The next American economy's learning series*. Roosevelt Institute. http://rooseveltinstitute.org/wp-content/uploads/2016/08/Lifelong-Learning-in-2040.pdf (accessed August 2017).

Sodhi MS and Son B-G (2010). Content Analysis of OR Job Advertisements to Infer Required Skills. *Journal of the Operational Research Society* **61:** 1315-1327.

Sodhi MS and Tang CS (2008). The OR/MS Ecosystem: Strengths, weaknesses, opportunities, and threats. *Operations Research* **56:** 267–277.

Sparck Jones K (1972). A Statistical Interpretation of Term Specificity and its Application in Retrieval. *Journal of Documentation*, **28:** 11-21

Sprague Jnr. R (1980). A Framework for the Development of Decision Support Systems. *MIS Quarterly* **4:** 1-26.

Stedman ME (1993). Statistical Pedagogy: Employing student-generated data sets in introductory statistics. *Psychological Reports* **72:** 1036-1038.

Stern DT (1998). Practicing what we preach? An analysis of the curriculum of values in medical education. *The American Journal of Medicine* **104:** 569-575.

Stonebraker M, Madden S, Abadi DJ, Harizopoulos S, Hachem N and Helland P (2007). The End of an Architectural Era: (It's time for a complete rewrite). *Proceedings of the 33rd International Conference on Very Large Data Bases*, 1150-1160.

Stork D (2003). Teaching Statistics with Student Survey Data: A pedagogical innovation in support of student learning. *Journal of Education for Business* **78:** 335-339.

Suh B, Hong L, Pirolli P and Chi E.H (2010). Want to be Retweeted? Large scale analytics on factors impacting retweet in Twitter network. In: *IEEE Second International Conference on Social Computing*, 2010. Minneapolis: IEEE, pp. 177-184.

Tanner Jr J F, Ahearne M, Leigh TW, Mason CH and Moncrief WC (2005). CRM in Sales-Intensive Organizations: A review and future directions. *Journal of Personal Selling & Sales Management* **25:** 169-180.

Taksa L (1992). Scientific Management: Technique or Cultural Ideology? *Journal of Industrial Relations* **34:** 365-395.

Tata S and Patel JM (2007). Estimating the Selectivity of tf-idf based Cosine Similarity Predicates. *SIGMOD Record* **36:** 75-80.

Tauer G, Nagi R and Sudit M (2013). The Graph Association Problem: Mathematical models and a Lagrangian heuristic. *Naval Research Logistics* **60:** 251-268.

Tavares LV (2002). A Review of the Contribution of Operational Research to Project Management. *European Journal of Operational Research* **136:** 1–18.

Taylor SJ (2018). Distributed Simulation: State-of-the-art and potential for operational research. *European Journal of Operational Research*. Forthcoming. https://www.sciencedirect.com/science/article/abs/pii/S0377221718303357 (accessed June 2018).

Teddlie C and Tashakkori A (2010). Overview of Contemporary Issues in Mixed Methods Research. In: Teddlie C and Tashakkori A (eds.). *Sage Handbook of Mixed Methods in Social & Behavioral Research*. Sage: California, pp 1-41.

Tempelaar DT, Rienties B and Giesbers B (2015). In Search for the Most Informative Data for Feedback Generation: Learning analytics in a data-rich context. *Computers in Human Behavior* **47:** 157-167.

Tezuka S, Murata H, Tanaka S and Yumae S (2005). Monte Carlo grid for financial risk management. *Future Generation Computer Systems* **21:** 811-821.

Thaler RH (2012). *Applause for the Numbers Machine*. The New York Times. http://www.nytimes.com/2012/11/18/business/how-pollsters-and-analysts-won-big-on-election-day.html (accessed March 2013).

Thorleuchter D, Van den Poel D and Prinzie A (2012). Analyzing Existing Customers' Websites to Improve the Customer Acquisition Process as well as the Profitability Prediction in B-to-B Marketing. *Expert Systems with Applications* **39:** 2597-2605.

Todd PA, McKeen JD and Gallupe RB (1995). The Evolution of IS Job Skills: A Content Analysis of IS Job Advertisements from 1970 to 1990. *MIS Quarterly* **19:** 1–27.

Tong S and Koller D (2001). Support Vector Machine Active Learning with Applications to Text Classification. *Journal of Machine Learning Research*, **2:** 45-66.

Trkman P, McCormack K, De Oliveira MPV and Ladeira MB (2010). The Impact of Business Analytics on Supply Chain Performance. *Decision Support Systems* **49:** 318-327.

Tukey JW (1977). Exploratory Data Analysis. *Addison-Wesley Series in Behavioral Science: Quantitative Methods*, Addison-Wesley: Reading, MA.

Turney P (2002). Thumbs Up or Thumbs Down? Semantic orientation applied to unsupervised classification of reviews. *Proceedings of the Association for Computational Linguistics,* pp. 417–424

Van der Aalst W and Damiani E (2015). Processes Meet Big Data: Connecting data science with process science. *IEEE Transactions on Services Computing* **8:** 810-819.

Van der Walt S, Colbert SC and Varoquaux G (2011). The NumPy Array: A structure for efficient numerical computation. *Computing in Science & Engineering* **13:** 22-30.

Varshney KR and Mojsilovic A (2011). Business Analytics Based on Financial Time Series. *IEEE Signal Processing Magazine* **28:** 83-93.

Vera-Baquero A, Colomo-Palacios R and Molloy O (2013). Business Process Analytics using a Big Data Approach. *IT Professional* **15:** 29-35.

Vere-Jones D (1995). The Coming of Age of Statistical Education. *International Statistical Review/Revue Internationale de Statistique* **63:** 3-23.

Vesset D, McDonough B, Morris H and Woodward A (2009). *Improving Organizational Decision-Making through Pervasive Business Intelligence*. http://www.idc.com/prodserv/pervasivebusintel.jsp (accessed April 2016).

Wagemans J, Elder JH, Kubovy M, Palmer SE, Peterson MA, Singh M and von der Heydt R (2012). A Century of Gestalt Psychology in Visual Perception: I. Perceptual grouping and figure-ground organization. *Psychological Bulletin* **138:** 1172-1217.

Wainwright MJ and Jordan MI (2008). Graphical Models, Exponential Families, and Variational Inference. *Foundations and Trends in Machine Learning* **1:** 1-305.

Walker M (2012). *Predictive, Descriptive, Prescriptive Analytics*. Analytic Bridge. http://www.analyticbridge.com/profiles/blogs/predictive-descriptive-prescriptive-analytics (accessed April 2016).

Waller MA and Fawcett SE (2013). Data Science, Predictive Analytics, and Big Data: A revolution that will transform supply chain design and management. *Journal of Business Logistics*, **34:** 77–84.

Wang X and McCallum A (2006). Topics over Time: A non-Markov continuous-time model of topical trends. In: *KDD '06,* August 20–23 2006, Philadelphia: ACM.

Ward SC (1989). Arguments for Constructively Simple Models. *Journal of the Operational Society* **40:** 141-153.

Waterson P (2011). World War II and other Historical Influences on the Formation of the Ergonomics Research Society. *Ergonomics* **54:** 1111-1129.

Wegman EJ, Carr DB, King RD, Miller JJ, Poston WL, Solka JL and Wallin J (1997). Statistical Software, Siftware and Astronomy. In: Badu GJ and Feigelson ED (eds.), *Statistical Challenges in Modern Astronomy II*. Springer-Verlag: New York, pp. 185-206.

Whiting R (2000). *MicroStrategy Boosts Speed, Accuracy*. Information Week. http://www.informationweek.com/793/micro.htm (accessed December 2012).

Winston WL (2004). *Operations Research: Applications and algorithms*, 4th edition. Brooks/Cole: Belmont, CA.

Wixom B, Ariyachandra T, Douglas D, Goul M, Gupta B, Iyer L, Kulkarni U, Mooney JG, Phillips-Wren G and Turetken O (2014). The Current State of Business Intelligence in Academia: The Arrival of big data. *Communications of the Association for Information Systems* 34: 1-13.

Wolpert D (1996). The Lack of *a priori* Distinctions between Learning Algorithms. *Neural Computation*, 1341-1390.

Woodford M (2010). Simple Analytics of the Government Expenditure Multiplier. *National Bureau of Economic Research* (No. w15714).

Woodruff MJ, Reed PM. and Simpson TW (2013). Many Objective Visual Analytics: Rethinking the design of complex engineered systems. *Structural and Multidisciplinary Optimization* 48: 201-219.

Working EJ (1927). What Do Statistical "Demand Curves" Show? *The Quarterly Journal of Economics*, **41:** 212–235.

Xiao B and Benbasat I (2007). E-Commerce Product Recommendation Agents: Use, characteristics, and impact. *MIS Quarterly* **31:** 137-209.

Xu DL, He W and Li S (2014). Internet of Things in Industries: A survey. *IEEE Transactions on Industrial Informatics* **10:** 2233-2243.

Yang J (2010). How we did it: Jie and Neeral on winning the first Kaggle-in-class competition at Stanford. http://blog.kaggle.com/2010/12/13/how-we-did-it-jie-and-neeral-on-winning-the-first-kaggle-in-class-competition-at-stanford/ (accessed March 2014).

Yuan ST and Sun J (2005). Ontology-based Structured Cosine Similarity in Document Summarization: With applications to mobile audio-based knowledge management. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, **35:** 1028-1040.

Zaki MJ and Ho CT (eds.), (2000). *Large-Scale Parallel Data Mining*. Springer: Berlin. Heidelberg: New York; Barcelona; Hong Kong; London; Milan; Paris; Singapore; Tokyo.

Zhang T (2004). Solving Large Scale Linear Prediction Problems using Stochastic Gradient Descent Algorithms. In: *Proceedings of the Twenty-First International Conference on Machine Learning* (p. 116).

Zhang J, Wang FY, Wang K, Lin WH, Xu X and Chen C (2011). Data-driven Intelligent Transportation Systems: A survey. *IEEE Transactions on Intelligent Transportation Systems* **12:** 1624-1639.

Zhang ML and Zhou ZH (2006). Multilabel Neural Networks with Applications to Functional Genomics and Text Categorization. *IEEE transactions on Knowledge and Data Engineering*, **18:** 1338-1351.

Zhao W, Ma H and He Q (2009). Parallel K-Means Clustering based on MapReduce. *IEEE International Conference on Cloud Computing*. Springer: Berlin, pp. 674-679.

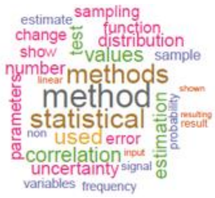
Item A Topic clouds generated in the computational literature review

Information Systems

Software

Security

Statistics



Video & Image Data

Learning analytics

Realtime analytics

Cloud Computing

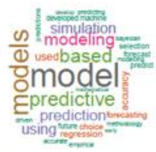
Health

Collaboration

Psychographics

Hierarchies

Predictive Models



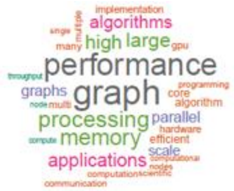
Literature Reviews

Social Media

The Web

Service Industries

HPC



Supply Chain

Knowledge

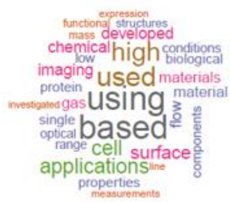
Big Data

Big Data Tools

Optimisation

Mobile & IoT

Physical Sciences



Questions & Objectives

Technology

Business & Management

Data Mining

Machine Learning

Decision making

Energy

Process & Products

Databases & Data

Visualisation

Volume & Scalability

Information

Science & Academia

Sports & Games

Politics & Culture

Importance & Growth

Finance

Text Data

Frameworks

Smart Cities

Item B Interview questions for potential employers

1) How do you define analytics in terms of your work?

- What types and forms of analytics are you involved in (predictive models, OR, etc.)?
- Are there areas of analytics you are seeking to introduce?
- What do you consider the impact of analytics (financial or otherwise) on your operations?
- How do you measure the success of analytics in your organisation?

2) What data storage and management tools (database management systems, Hadoop, NoSQL, etc.) do you use?

- How are these used in the operations and what impact do you think they have?
- If you use both traditional database tools and NoSQL, how are they combined?
- Are you looking for new employees to have any experience with specific tools or systems?

3) What analysis tools (e.g. SAS, R, SPSS, Matlab, Excel, etc.) do you use?

- Which of these are most important and what purposes are they used for?
- Are you looking for new employees to have any experience with specific tools or languages?

4) Are you looking for employees to have higher-level programming or software development skills?

- If so, which specific languages are most important (e.g. C++, Java, Python, PHP, etc.)?

5) Do you use any specific data visualisation tools or methods?

- What experience in visualisation are you looking for new employees to have?

6) How is analytics managed within your operations?

- Is there are a single department or is it cross-functional?
- Are analysts and IT functions located in the same team? How do team members from each work together?
- Do frontline staff perform their own analyses? How is this enabled?
- To what extent are you looking for new employees to have specialisms in all aspects of analytics (technology, quantitative methods, etc.)?

7) Have you been recruiting or are you currently recruiting for analytics staff?

- If NO – do you intend to?
- If YES – how difficult is/was it to fill these roles satisfactorily?
- What roles do you particularly seek to fill? Which departments or specialisms?

- Do you hire fresh graduates (bachelors or post-graduate)?
    - If YES – what specialisms and disciplines do you recruit from?

8) Are you or your organisation involved in initiatives to support the training of new graduates (guest lectures, case study materials, etc.)?

- If YES – what form does this take?
- If NO – is this something you may consider in the future?

9) What are the key skills, capabilities and experiences you think universities should be delivering to graduates?

10) How do you think analytics will develop in your organisation?

- Do you think it will become more important? Do you foresee future investments?
- If YES – what form do you expect this to take (personnel, technologies, methods, etc.)?
- If NO – what are the reasons for this?

11) How do you think analytics will develop in general?

- Are there any areas you expect to see innovations in?
- Are there areas where current analytical capabilities are failing or can be improved?

Item C Interview questions for academics

1) Can you give me a little information about yourself and your school/university?

- And your involvement in analytics or data science?
- And your involvement in OR (or other subjects)?

2) What differences do you see between analytics and OR (or other subjects)?

3) What academic traditions do you think analytics draws from?

4) Do you offer analytics courses?

- If NO - Are there plans to do?
- If NO - Is there any internal or external demand to do so?
- If NO - What extent do current offerings fit in this area?
- If YES - Masters/Bachelors, what school/faculty?
- If YES - What demand is there? How does this compare?
- If YES - What differences do you consider there to be between your analytics course and other courses you run (e.g. OR)?

5) What are the core skills that you think analytics courses should teach?

6) What are the core skills that you think OR courses should teach?

7) To what extent do you think that core analytics skills are delivered in OR degrees (or other subjects)? What is missing?

8) Are there programming languages that should be taught in analytics or OR courses? Do they differ?

9) Is there specific software?

10) What forms of datasets should be used - and where can they be sourced? Is there a difference between OR and analytics in this respect?

11) What types of problems and exercises should be presented? Is there a difference between OR and analytics in this respect?

12) What value do you place on internships or consultancy projects?

13) What barriers do you see that complicate the creation of analytics degrees?

- Availability of experienced teaching staff?
- Availability of data and tools?
- Time?

- Different schools teaching the relevant skills?

14) How do you think analytics and OR degrees will develop in general?

15) To what extent do you think that universities need to adapt to current business trends and how much do they need to maintain the academic traditions of disciplines?

Item D Topics and most likely words from the job advert analysis

Topic Label	Term #1	Term #2	Term #3	Term #4	Term #5	Term #6
Process monitoring	qualiti	process	ensur	complianc	manag	control
Forecasting	price	retail	forecast	categori	demand	plan
Software (development)	engin	develop	experi	softwar	team	technic
Digital marketing	search	googl	bing	keyword	dell	engin
Analysis (quantitative)	data	report	analysi	analyst	statist	perform
Customer support	custom	servic	support	provid	issu	call
Tax and audit	tax	bull	touch	fas	llpdeloitt	firm
Language and culture	english	fluent	spoken	command	coursework	china
Cloud and NoSQL	cloud	virtual	aml	infrastructur	vmware	directori
Analysis (business)	busi	requir	develop	system	process	design
Employment (other)	hse	comcast	kaiser	amend	permanent	rental
Clinical	clinic	studi	regulatori	trial	pharmaceut	statist
Employment (other)	microsoft	sharepoint	window	rsquo	weather	block
Employment (other)	gender	year	experi	least	employ	ibm
Marketing	busi	market	team	experi	strategi	develop
Scientific	scientif	commerci	biostatist	optimis	des	behaviour
Project management	project	manag	team	plan	experi	resourc
Employment (other)	train	program	requir	work	provid	assist
Employment (other)	branch	graph	guest	car	flight	lieu
Employment (other)	region	local	canada	beverag	canadian	label
Employment (other)	status	employ	disabl	protect	veteran	nation
Software (quality & testing)	test	qualiti	case	plan	valid	ute
Visualisation & interfaces	design	user	experi	interact	inspect	usabl
Programming	develop	experi	applic	web	design	softwar
Security	secur	inform	protect	crimin	network	access
CRM	figur	appl	crm	salesforcecom	defend	hoop
Financial (investment and trading)	financi	invest	bank	manag	trade	market
Mobile and games	autom	akamai	anywher	android	spc	exp
Big data	data	experi	build	big	scale	system
Employment (other)	cancer	los	sandia	nuclear	angel	biomed
Employment (other)	san	tackl	built	francisco	surround	ceo
Intelligence & operations	children	defens	simul	air	mission	oper
Financial (analysis)	capit	one	actuari	sas	year	job
Management (teams)	manag	develop	team	plan	respons	experi
Management (skills)	busi	manag	develop	team	process	plan
Advertising	advertis	media	campaign	video	onlin	digit
Database	databas	sql	oracl	experi	server	develop
Employment (other)	global	countri	region	interdisciplinar	worldwid	telecom
Sales	sale	account	custom	busi	manag	new
Chemistry	affair	chemistri	intel	semiconductoı	parttim	basketbal
Employment (other)	end	user	peoplesoft	versatil	front	workday
Employment (other)	edg	cut	india	nonprofit	toler	movement
Manufacturing and SCM	suppli	manufactur	chain	process	oper	sap
Employment (other)	resum	pleas	appli	email	posit	inc
Real estate	real	estat	green	programmat	utabl	exhaust
Employment (other)	survey	compens	child	incent	hris	postgradu
Employment (other)	must	clearanc	oper	nation	engin	secur
Public sector (governing)	state	feder	govern	servic	agenc	includ

Item D (cont.)

Topic Label	Term #1	Term #2	Term #3	Term #4	Term #5	Term #6
Employment (other)	youll	street	mark	smarter	liter	climat
Employment (other)	work	game	make	peopl	world	like
Solutions and architecture	solut	technolog	architectur	enterpris	technic	design
Marketing campaigns	market	campaign	manag	event	brand	communic
Product development	product	manag	new	featur	launch	releas
Employment (other)	inform	bloomberg	parti	compani	third	applic
Financial (control)	financi	account	report	financ	prepar	manag
Employment (other)	insur	membership	compani	fit	union	properti
Employment (other)	america	north	europ	asia	nutrit	seminar
Employment (other)	screen	check	employ	drug	background	workplac
Other	sociolog	cisco	invent	rout	transit	voic
Modelling	statist	data	analyt	model	analysi	experi
Employment (other)	will	team	work	look	can	need
Communication skills	skill	abil	work	experi	strong	communic
Human resources	human	resourc	recruit	employe	psycholog	talent
Employment (other)	vehicl	young	automot	driver	licens	dealer
Financial (audit)	audit	risk	control	intern	assess	manag
Employment (other)	deloitt	consult	servic	busi	llp	subsidiari
Public sector (services)	patient	famili	hospit	treatment	communiti	registr
Employment (other)	will	role	work	team	within	skill
Networking (computing)	network	storag	knowledg	protocol	engin	infrastructur
Employment (other)	forc	prescrib	comment	inperson	entitl	empathi
Software (use)	broadcast	whenev	par	tabil	spotfir	roleth
Employment (other)	hour	week	offic	assist	day	schedul
Employment (other)	digit	social	consum	media	brand	sport
Publishing	content	publish	write	web	websit	edit
Employment (other)	benefit	compani	offer	competit	includ	opportun
Other	leur	une	vous	export	surgic	qui
Employment (other)	water	carrier	micron	onthejob	exclus	phoenix
Travel (sector)	hotel	travel	european	book	compani	emea
Medical	physician	popul	biolog	medicin	therapi	diseas
Employment (other)	other	chang	work	general	use	member
Consulting	client	consult	servic	manag	solut	industri
Employment (other)	health	care	clinic	medic	healthcar	provid
Employment (other)	school	internship	teach	cours	date	month
Employment (other)	mobil	devic	wireless	companywid	phone	authent
Milatary	militari	threat	airlin	concur	aviat	termin
Engineering & safety	safeti	engin	energi	equip	electr	environment
Employment (other)	programm	healthcar	registri	cycl	type	mcgladrey
Employment (other)	work	compani	opportun	peopl	world	help
Systems management	system	support	experi	oper	manag	applic
Research	research	conduct	studi	qualit	secondari	quantit
Employment (other)	visa	sponsorship	northern	societi	sme	author
Social media	social	facebook	media	youtub	communiti	linkedin
Financial (credit)	risk	credit	card	model	manag	bank
Machine learning	learn	machin	comput	algorithm	program	scienc
Business intelligence	data	intellig	warehous	busi	teradata	model
Employment (other)	servic	center	deliveri	desk	level	citi
Employment (other)	psycholog	california	cognit	boston	walk	fellow
Ecommerce	onlin	web	site	ecommerc	optim	websit
Employment (other)	requir	perform	duti	must	job	abil
Employment (other)	will	student	program	univers	posit	candid

Item E Categorisation of interview participants to 'approaches'

Interviewee	Assignment	Reason
Analytics Manager (Telecoms)	Operationalisation	Team included two "computer scientists" how would help with database work particularly. Most models later integrated into operational systems.
Government Analytics Manager	On Demand	No technology-orientated team members. Resources from other parts of Government could be used.
Analytics Consultant (Smaller Management Consultancy)	On Demand	Primarily focused on *quantitative methods* and *decision making*. Other *technology* resources could be sourced elsewhere in the organisation.
Analytics Manager (Health)	On Demand	No technology-orientated team members. Resources from other parts of Government could be used.
Analytics Manager (Utilities)	On Demand	No permanent technology-orientated staff. Worked with other departments for such resources.
Analytics Consultant (Finance)	Unicorn / Operationalisation	Staff involved in activities from across the analytics spectrum. Most models were later operationalised by other teams.
Analytics Manager (Online Travel)	Unicorn / Operationalisation	Tried to recruit staff who had awareness (at least) of the full stack of analytics work. Most analytics would be integrated into the website and other systems.
Analytics Manager (Public)	N/A	Virtually all projects were ad-hoc so little call for databases or other *technology* aspects of analytics.
Marketing Analytics Consultant	On Demand	Primarily focused on *quantitative methods* and *decision making*. Other *technology* resources could be sourced elsewhere in the organisation.
Software Vendor (Data Management)	Unicorn / Operationalisation	Most staff working across the spectrum of analytics, although with specialisations. Internal work mostly updated into the software.
Government Analytics Manager (Finance)	On Demand	No technology-orientated team members. Resources from other parts of Government could be used.
Software Vendor (Analytics General)	On Demand / Operationalisation	Some staff working in a consultancy capacity, using IT rescources on demand. Other work towards the software was later operationalised.
Analytics Manager (Energy)	On Demand	Team was essentially comprised of two teams, one *technology* team; one completing *quantitative methods* and *decision making* tasks. Resources shared as required.
Analytics Manager & Analytics Consultant (Larger Management Consultancy)	On Demand	Team was essentially comprised of two teams, one *technology* team; one completing *quantitative methods* and *decision making* tasks. Resources shared as required.
Software Consultant (Simulation - Processes)	Operationalisation	Various team members with different skills and roles across the analytics spectrum. Most analytics operationalised into their software.
Marketing Analytics Manager	Technical + Business	A business facing team (charged with consultancy-type tasks) and a team managing technologies and modelling activities.
Software Vendor (Simulation - All)	Unicorn / Operationalisation	Most staff involved with all aspects of analytics. Much of the work is ultimately included in the company's software.

Note: Some cases excluded for one of two reasons. (1) They are already included in the matrix analysis (section 6.4); (2) their organisation is not appropriate to include, such as recruitment consultants.

Item F A Brief Introduction to Text Analytics

This final appendix item is designed to give a general introduction to the text analytics field. The motivation is two-fold. Firstly, this allows us to expanded upon some of the some of the methods utilised in the thesis (without adding unnecessary bloat to the main flow of the work). Secondly, with such methods representing a key part of our methodology, and with the usage of such methods relative rare in work of this kind and in this specifc area of research, this discussion can help add flavour and further insight to this relatively new area of study.

This introduction is arranged as follows. Firstly, some of the specific challenges associated with text data will be discussed. This is followed by a description of the most common methods of pre-processing and cleaning the data for use in analytical models. This introduction concludes by presenting three common 'classes' of applications in this space: *descriptive* (for instance, for understanding the themes and topics common to a document); *predictive* (classification and regression-type applications); and *comprehension* (for example, to be used in "chatbots" or other forms of artifical intelligence).

Challenges of Text Data

Text is often considered, in respect to the commonly three V's of big data representation (Laney, 2001) as "unstructured data" (variety). Whilst "unstructured" is in many ways a poor descriptor, as text data clearly presentes structures (in terms of documents, paragraphs, sentences and the words themselves), it is fair to say that these structures can be challenging to work with, and without some of the consitency and familiarity of working with more traditional data, such as financial records.

In qunatitative analyses, the most obvious first issue is that text data is by default non-numeric. However, in many cases counts of specific elements are the key concern in text analytics; most commonly word counts, but, depending on the application we may also consider number of words in a sentence (to measure sentence-complexity), the number characters in each word (to measure language-complexity), or similar. In comprenshion-type tasks, the ordering of items (words) is often crucial, which obviously has a numerical representation in terms of word position within a sentence.

In applications based on word counts, key to many text analytics approaches, another key difference is the distribution of word counts. It is rare that these follow a normal distribution, with a small number of words occuring in high frequency (e.g. "I", "the" and "and" itself), while the majority of words occur with low frequency (particularly when comparing a collection of documents where the majority of words in a vocabularly will have zero frequency in most documents).

Additionally, documents will typically contain non-text characters. These include punctuation, numbers, and, specifically when working with machine-generated content, other characters such as special characters or icons. If the documents analysed come from different locales, or are scientific documents, an additional challenge can be managing characters from different alphabets, from the greek letters used in statistical notation, to umlauts and accents applied to characters in some langauges.

One of the biggest challenges is the inherent flexibility and variety of language use. Whereas numbers carry precise and fixed meanings (100 is always 100), words can be used in multiple ways and to denoted different concepts, even within the same sentence. Some key concepts here are synonomy (where multiple words can have the same or similar meaning), polysemy (where words can have multiple meanings), and hyponymy/hypernymy (where words share semantic categories – for instance, "apple" and "orange" are co-hyponyms of "fruit" (their hyerpnym). While challenging, such elements are mostly "knowable" in the sense that there are rules that govern these realtions. However, when dealing with non-technical texts, more complex and challenging issues arrise such as the use of slang, short-hand abbreviations (such as "pls" instead of "please") or irony and sarcasm.

There are also practical concerns. In many cases, if not the majoirty of cases, the data used for text analytics applications will be sourced or streamed from the internet. This necessitates often significant cleaning requirements such as parsing HTML elements, removing scripts (e.g. JavaScript code), and/or dealing with URLs and path directories. In the case of streamed data, many of these processes need to occur in near real-time.

In summary, while text does have inherent structures and numerical properties, which facilitate many opportunities for analytics to be performed, there are clear challenges and issues not present when dealing with "traditional", numerical-type data. However, many of these challenges can be met through the application of well-established pre-processing and cleaning techniques, some of which will be discussed in the next section.

Processing and Cleaning Text Data

Whilst the challenges discussed in the previous section are often non-trivial, there are solutions to them. In most text analytics applications the series of required steps can be formalised into an algorithm and performed automatically as part of a data pipeline. Some typical examples of such steps are presented in this section.

In some cases these steps are reasonably obvious and standardised. For example, parsing HTML, converting characters to lower case (as upper- and lower-case characters are read as distinct by computers), and, depending on the task, removing non-printable characters, punctuation, numbers, and/or URLs. In other cases this decision may be more subjective and contentious.

One example is the inclusion, or otherwise, of 'stopwords'; short, very frequent words such as prepositions that tend to have limited information in analytics applications. By their very nature, what constitutes a stopword is by nature contentious and task specific. In other cases, the exclusion of stopwords would be highly determinental to the efficacy of the application. For instance, computational sylometry, the automated identification of a text's author (useful for detecting plagiarism or resolving authorship disputes), the use of what would be considered stopwords in other applications, can provide a huge amount of information to the document's authorship, as often these words are used in a more subconcious way, and not uniformly between different authors.

Whether the ordering of words is retained is another key concern, and one which is highly dependent on the task in hand. In many applications, particularly those based on word counts, the ordering of words is often removed, such that a document becomes a 'bag-of-words' rather than a series of sentences. Most commonly this means creating a document-term matrix (DTM) whereby each document represents a row, the column represent each word in the vocabulary, and the individual elements are the per-document, per-word frequencies. Such a representation has clear benefit for tasks based on summarising, comparing or making predicitons about documents, most notably because this becomes a numerical representation of a document as a distribution of words. However, if the task is to understand meaning of requests or short-form statements (for instance, a chatbot or a FAQ (frequently asked questions) section of a webiste) then the ordering of words can be critical. Represented in a DTM, with standard stopwords removed, the phrases "research on business operations" and "operations research in business" would be read as the same.

A similar concern regards how tenses and similar variations in words are treated. In common speech, "managing" and "managed" are clearly the same concept, but if computing a DTM these are unique entities. Stemming provides a solution to this, whereby all the words in a corpus (collection of documents) are reduced to their shortest stem. In the above example, both terms would be reduced to "manag", and therefore treated as a single entity in the DTM. Obviously "manag" is not a real word, so an alternative is lemmatisation where words are reduced to their shortest lemma (real word) rather than stem; meaning in the above example "manage" would be used. Whilst this is intuitively more satisfactory, there is limitation in that each transformation necessitates dictionary lookups, which can have considerable computational cost. Although stemming and lemmatisation have obvious benefit in frequency-based tasks, this is not without cost. Although “managing” and “managed” are functionally the same word, the meaning differs slightly from “manager” or “management”, both of which would be reduced to the same stem/lemma.

The larger, more variable tasks such as dealing with polysemy, slang or irony, are typically more complex and less standardised. Often the solutions depend on whether the approach is comprehension-based (where ordering has been retained and stemming/lemmatisation is unlikely to be used), or frequency-based (e.g. data transformed to a DTM, and often stemmed). In the case of ther former, the positioning of words in sentences provides some evidence. As a toy example, “I’m happy” and “I’m *not* happy” changes meaning completely with the positioning of the word “not”. For polysemy, the meaning of other words in the sentence or in previous sentences can help determine the likely meaning of a given term. In frequency-based analysis, particularly where DTMs and/or stemming are used, this becomes considerably more problematic as surrounding words are removed. There are methods that can help meet these challenges, but the task is considerably more complicated.

Descriptive Approaches

The first category of algorithms and approaches that will be discussed, are those that seek to describe datasets. While descriptive analytics is often described as the “lowest” form of analytics (see section 1.2), many descriptive apporaches to analysing text data are comparatively complex, and potentially very powerful. For instance, a company such as Google, that indexes websites to match user search queries, is dependent on such methods to perform these tasks.

One of the best known of these approaches is sentiment analysis. Although there are precursors in the literature, work in this area accelerated in the early 21st Century, coinciding with Web 2.0 and the prolification of user generated content, social media and online reviews (e.g. Pang *et* al, 2002; Turney, 2002). Sentiment analysis varies in complexity, and typically therefore accuracy, from lookups to generic word-lists that have an associated polarity (degree of positivity or negativity, for instance, on a scale from -1 to +1), to custom built analyses where polarity is learned from the data (for instance, by building a classification or regression algorithm based on partially labelled documents).

Although sentiment analysis can have value, it is relatively reductionist in its analysis of text, as no meaning is actually derived, just a single measure of polarity. In many cases, the subject matter of texts are of more importance. At the more basic level, this can come through finding the aspects of different documents that are different, or by finding other documents which are similar to it. Similarity in text is a compartively well-established field of research. A variety of methods are available for such tasks, mostly based on mapping space between two documents, such as the Jaccard similarity co-efficient (e.g. Niwattanakul *et al*, 2013) or cosine similarity (e.g. Yuan and Sun, 2005). Alternatively, methods can be used to find the words that offer the most disriminatory power between documents, for instance using the term frequency – inverse document frequency (TF-IDF) algorithm (Sparck Jones, 1972). This algorithm can be used to transform documents so that words are scored by their frequency (TF) adjusting for their relative frequency in other documents in a corpus (IDF). Uncommon words that are infrequent in a given document score low on the basis of frequency; while words that are frequent in a document, but are also frequent elsewhere (such as "the" or "and") are penalised by this and also scored low. The words that score highly are those that are frequent in one document, but infrequent in others, and therefore demonstrating the relative importance of that word to the document. Such as transformation is relevant to document similarity tasks, by allowing the researcher to match based on the terms that score highest in TD-IDF (and are therefore relatively unique to these two documents).

Extending this approach, methods are presented that move beyond single-word matching to groupings of words into components. This family of methods is typically known as topic models (which have been used in both the computational literature review of section 1.3, and the job adverts analysis of chapter four). The first iteration of topic models were based on singular value decomposition (SVD), a well established approach to matrix decomposition. Latent semantic analysis (LSA) is the most famous of these, effectively principal component analysis (PCA) for text data (Dumais, 2004). However, as with PCA, the issue with such an approach is that it cannot deal with synonymy. Effectively each unique data point (i.e. each unique word) takes a single position in the co-oridnate system of the transformed dataset, so that effectively every instance of a word has the same "meaning". As a more layman's example, "lead" as in "sales lead" would have the same meaning as the chemical element "Pb". Obviously this is both theoretically dissatisfactory and also practically problematic.

Accordingly, alternative solutions were developed that could counter this principal disadvantge of LSA. Firstly, probabalistic latent semantic indexing (pLSI) was presented in Hofmann (1999). The advantage of pLSI is that rather than decomposition of the word frequency tables, which cannot accommodate the assignment of the same word to multiple compotents, pLSI is a mixture decomposition of a latent variable model for word co-occurrence in the documents. In other words, word-assingment to a given documents is determined by latent factors/topics (and therefore estimated probabalistically, usually via the EM (expectation-maximization) algorithm) analagous to the subject matter of the document. In doing so, words can effectively take multiple "positions" (theoretically one per topic), each of which can have a different semantic meaning, effectively bypassing the problem of polysemy.

While pLSI offers a marked improvement over its predecessor, and remains a widely-used solution to topic modelling, it too is not without issue. In particular, although it represents a generative model of its training set, the input corpus, it cannot be used as a generative model for future data (i.e. the model cannot be used to assign probabilities to new, unseen documents), and secondly there are risk of over-fitting as paramters grow linearly according to the size of the corpus (Blei *et al*, 2003).

Latent Dirichlet allocation (LDA), described in section 2.4.1, provides an alternative which directly addresses these issues (Blei *et al*, 2003). In this instance, rather than fixing the algorithm to an index associated with each document, a Dirichlet prior is used which can also be applied to unseen documents, thus making LDA a generative model for both seen and unseen data. By the same token, the number of priors is fixed in size to the number of topics not to the number of documents, so it is no longer tied to the corpus size, limiting the potential for over-fitting.

Empricial evidence suggests that in many practical scenarios, and for wholey unsupervised tasks without any required addition of unseen documents, there is little significant difference between pLSI and LDA in terms of quality of results (e.g. Masada *et al*, 2008). However, in that the model is more theoretically appealing and presents the potential for use in estimating new documents, LDA has seemingly grown to be the most widely used of the two, based upon annecodotal observation of applied publications in this space.

There are multiple extensions of the LDA algorithm in the literature, not least the Online approach of Hoffman *et al.* (2010) used in this thesis to analyse a larger corpus of job adverts than would have been practical with the standard algorithm (in the sense of both time and required processing power). Additional noteworthy additions include:

- An extension for supervised tasks such as predicting star ratings of reviews based on their text (e.g. McAuliffe and Blei, 2008);
- Hierachical approaches whereby topics can represent 'parent nodes' for other topics (e.g. Griffiths *et al*, 2004). For example, an "animals" topic may produce 'child nodes' of topics on "cats and "dogs";
- Topic models with a time-based element, such that topics can develop, and be tracked, over time (e.g. Wang and McCallum, 2006).
- Topic models with co-variates. Effectively using meta-data to help define the topic model strucure (e.g. Roberts *et al*, 2014).

Predictive Approaches

Another common task in text analytics, and indeed the whole of machine learning, is prediciton. As an example, an ecommerce company may want to make predictions about which products to recommend to users based on analysis of the textual description of consumer reviews. As with much of machine learning (certainly supervised or semi-supervised machine learning), typically this is in the form of classification (where categories or labels are predicted), or regression (where continuous values are predicted).

Also, almost all of the more common prediction algorithms used in machine learning are, and have been, applied to text data, particularly if transformations to numerical values have been applied (for instance, into a document-term matrix). As demonstration, table 37 highlights examples of the more common algorithms from the literature.

Table 37 Examples of prediction algorithms used on text data

Algorithm	Application	Examples
Decision trees / random forests	Both	Schmid (1994); Lior (2014)
Generalised linear models	Regression*	Genkin (2007); Joshi *et al.* (2010)
Gradient descent	Both	Zhang (2004); Shahnaz et al. (2006)
Naïve Bayes	Classification	Frank and Bouckaert (2006); Chen *et al.* (2009)
Nearest neighbour	Both	Cheung and Fu (1998); Davy and Luz (2007)
Neural networks / deep learning	Both	Zhang and Zhou (2006); Venugopalan *et al*. (2014)
Support vector machines (SVM)	Both	Joachims (1998); Tong and Koller (2001)

**Although linear models are mostly applied to regression problems, it is possible to use methods such as logistic regression for classification*

Additionally, it is possible to use such algorithms in combination. Examples include boosting (where multiple models are applied sequentially to optimise the error of previous models), bagging (different models 'voting' for the class assignment in classification problems), or 'stacked' algorithms (where multiple models are used to make initial predictions, and the results of these are fed into an algorithm as input for the final prediction).

Comprehension Approaches

The final category analysed here are algorithms and methods designed to comprehend the specific meaning of text inputs. Examples of applications here range from query processing (such as in search engines) to conversational artificial intelligence (such as chatbots on websites).

Unlike many of the previously discussed approaches, tasks such as these often involve different pre-processing steps (as discussed in the "Processing and Cleaning Text Data" section). For example, while transformation to a document-term matrix (the 'bag-of-words' assumption) is common to many of the previously discussed approaches, losing the ordering of words in a sentence can present significant issues in comprehension tasks.

Indeed, often comprehension approaches seek to model the ordering of words. Whereas other approaches will typically focus on unigrams (single words), in many comprehension tasks it becomes necessary to consider words that occur together as single entities; such as the words "data" and "science", if sequenced in this order, as "data science". This can go beyond bigrams (two-word pairs) to any number of combined words (n-grams).

Additionally, analysis of sentence structure can be used, for example, to form part-of-speech tagging or parse trees. In applications of the latter, a sentence will be split into phrases, such as a noun phrase and a verb phrase, so that the interaction between them can be identified (for instance, and to continue the example, a verb phrase can indicate a requested action requested towards the noun in the noun phrase).

Such approaches are well established in the literature, with its roots in linguistic studies from the 18^{th} Century or earlier (e.g. Robins, 1997). However, the area has seen considerable advancement in recent years due to the growth in deep learning methods. In particular, recurrent neural networks (RNNs) such as long short-term memory (Hochreiter and Schmidhuber, 1997) have been used in conversational artificial intelligence applications such as Amazon Alexa (Mass, 2018). Although Alexa works with audio input, the system effectively performs the same function as a chatbot processing text data.

The principal difference between recurrent neural networks and 'vanilla' artificial neural networks is that multiple inputs can be processed in the algorithms memory (for a short period of time), rather than singular inputs feeding forward (or backward) in the network. The importance this has for conversational approaches is that it allows the algorithm to process a series of connected inputs (e.g. sentences) where meaning from an earlier input can be imbued into a later one. However, it is also necessary that the algorithm can 'move on' from this at given times. Consequently, the long short-term memory (LSTM) algorithm uses "forget gates" to effectively terminate the 'memory' of previous inputs.

For instance, consider the input: *"Pre-processing data is different in conversational approaches. Often word order is maintained. Analytics is widely used in many organisations"*. To fully understand the meaning of the second sentence it is necessary to have the context of the first 'in-memory'. However, the third sentence has no such requirement, so a well-tuned algorithm can 'forget' the previous inputs to process it.

Summary

Over the course of this item, many of the most common approaches to text analytics are presented in order to provide a gentle introduction to the area. In doing so, the hope is that extra context has been provided to some of the methods in this thesis, and information given on this growing area for the interested reader.