DEMANDED AND IMPARTED BIG DATA COMPETENCES: TOWARDS AN INTEGRATIVE ANALYSIS

Matthias Murawski
ESCP Europe Business School, Berlin, Germany, mmurawski@escpeurope.eu

Markus Bick
ESCP Europe Business School, Berlin, Germany, Bickmbick@escpeurope.eu

Follow this and additional works at: http://aisel.aisnet.org/ecis2017_rp

Recommended Citation
http://aisel.aisnet.org/ecis2017_rp/89
DEManded and Imparted Big Data Competences: Towards an Integrative Analysis

Research paper

Murawski, Matthias, ESCP Europe Business School, Berlin, Germany, mmurawski@escpeurope.eu
Bick, Markus, ESCP Europe Business School, Berlin, Germany, mbick@escpeurope.eu

Abstract

Exploiting big data seems to be an important success factor for companies in the digital age. However, recent studies show that there is a short supply of professionals who are able to deal with data appropriately. This is at least partly caused by a mismatch between university offerings and presumed industry needs. This study analyses two related questions. First, what competences are actually required for being a data professional? Second, what competences are imparted through data-related master’s programmes? These questions are answered by applying a topic model approach (first question) and deductive content analysis (second question). By using the same set of competence dimensions, the answers to these questions are used to discuss the overall issue of how curricula are aligned with workforce demands for data-related competences. The focus is placed on the UK market that suffers from a shortage of data professionals particularly in the financial industry. We find that companies require ‘all-rounders’ who possess strong technical, analytical, and business competences, while master’s programmes rarely impart business competences. Main contributions include an empirically derived typology of data professionals, the application of a topic model for IS research, and an analysis framework that allows universities to critically assess their offerings.

Keywords: Data Professionals, Integrative Analysis, Job Advertisement Analysis, Master’s Programme Analysis, Topic Model.

1 Introduction

How can companies be and remain successful in an increasingly digitized world? While a detailed answer always depends on the individual case, companies such as Facebook, Amazon, or Alphabet show that one of the general keys to success is the exploitation of big data (Chen et al., 2012; LaValle et al., 2011; Lycett, 2013; Sharma et al., 2014). Specific attributes used to define big data are the ‘4Vs’: volume, velocity, variety, and value (Lycett, 2013). Data analytics, defined by the Association for Operations Management as ‘a collection of data and technology that accesses, integrates, and reports all available data by filtering, correlating, and reporting insights not attainable with past data technologies’ (N.N., 2012) potentially lead to insights that are able to guide both future strategies and day-to-day operations of a firm (Lycett, 2013). A growing number of companies have realised this trend and aim to generate value based on big data (LaValle et al., 2011; Provost and Fawcett, 2013). However, the best data is worthless if there are no data professionals who are able to make them valuable (Chen et al., 2012; Tambe, 2014). Being able to handle data (e.g., selecting, analysing, and interpreting [big] data) is an essential digital competence area (McAfee and Brynjolfsson, 2012), but required individual competences are rarely investigated. Besides some exceptions (Debortoli et al.,
a profound and empirically derived set of data-related competences is missing. Although data-related competences are relevant for many occupations, this paper focuses on business data specialists, meaning those professionals whose primary job is to deal with big (business) data. When screening the job market, various job titles for this group of people exist (e.g., data scientist, data analyst, business intelligence analyst, or digital analyst). There are initial attempts to define boundaries of certain occupations such as data scientists (Provost and Fawcett, 2013), but the above mentioned terms seem to be used interchangeably especially in practice, like in job advertisements. In this paper, we use the term data professionals to cover all jobs focussing on dealing with big (business) data.

Considering the job market, many companies have an increasing demand for data professionals, but there is a shortage of professionals available (Dubey and Gunasekaran, 2015; Manyika et al., 2011; Wixom et al., 2014; Wixom et al., 2011). This results in high salaries and attractive career opportunities (N.N., 2016). Many open positions cannot be filled; this might lead to negative consequences for demanding companies since they are unable to exploit and analyse their data properly. Academia has realised this fact and started offering data-related master’s programmes; however, these programmes are often not aligned with presumed industry needs (Turel and Kapoor, 2016; Wixom et al., 2014; Wixom et al., 2011). In line with these developments, a new wave of research related to data-focussed master’s programmes and their curricula can be observed. Amongst others, this includes ‘calls to action’ for universities regarding the necessity to respond to emerging data-related market needs (Dubey and Gunasekaran, 2015; Wixom et al., 2014; Wixom et al., 2011), suggestions for curricula design (Gupta et al., 2015; Mitri and Palocsay, 2015; Schoenherr and Speier-Pero, 2015), and maturity assessments of business schools regarding their data courses offerings (Turel and Kapoor, 2016).

The main question of this paper is how curricula are aligned with workforce demands for data-related competences, thereby going a step beyond existing literature by analysing both perspectives—demanded competences and imparted competences—through an integrative approach. For this reason, the information systems (IS) competence framework by Todd et al. (1995) has been identified to provide comparison dimensions. Although the topic at hand seems to have strong relations to other disciplines such as human resources or educational sciences, the IS field holds a leading position in big data research in general (Agarwal and Dhar, 2014) as well as in organizing educational data-related programmes, as it comprises both business as well as technical elements (Chen et al., 2012; Mitri and Palocsay, 2015; Wixom et al., 2011). Regarding the geographic context, existing studies mainly deal with the United States (US). The underlying rationale is that scholars consider the US benefit from a great number of both IS/IT jobs and study programmes that can be analysed. We believe that it is worth considering other market regions as well. Thus, in this study, our approach is applied to the United Kingdom (UK). The UK shows a high level of technology-related jobs, especially in the financial sector that is characterised by an already extensive but still increasing use of big data analytics (Brown et al., 2011; LaValle et al., 2011). Similar to the US context, a substantial shortage of data professionals hinders British companies in leveraging the value of big data (Harris and Eitel-Porter, 2015; Harris et al., 2013), while future demand for data professionals is expected to greatly increase according to an industry report (N.N., 2014). Furthermore, a large number of well-established universities offering data-related master’s programmes that are necessary for our analysis can be ensured. Based on this, the following set of research questions is formulated:
RQ1: What are required competences of data professionals in the UK?

RQ2: What competences are imparted by data-related master’s programmes in the UK?

RQ3: How do data-related master’s programmes match competence requirements for data professionals in the UK?

The remainder of this paper is as follows. In section 2, existing literature dealing with required competences of data professionals, data-related master’s programmes, and IS competence frameworks are presented. Section 3 contains both the explanation and application of our research approach, including a job advertisement analysis and a master’s programme analysis. Both are investigated by applying the same competence dimensions, which allows the merging and discussion of results in section 4. The paper ends with a conclusion and suggested avenues for further research in section 5.

2 Related Work

This chapter presents the current state of research regarding both required competences of data professionals and data-related master’s programmes. For comparing required and imparted competences, a consistent set of competences is necessary. We have selected the IS competence framework of Todd et al. (1995) that provides broad but suitable categories for digital competences and is successfully applied in similar studies (e.g., Müller et al., 2014).

The notions of competences, work-related knowledge, skills, and abilities are often used interchangeably in research articles (Müller et al., 2014). In this study, we use the term competence to mean a combination of abilities, (work-related) knowledge, and skills held by an individual (Nordhaug, 1993). Abilities innately belong to an individual (like the ability to engage in logical reasoning). Knowledge means a theoretical understanding of a concept, while skills are the practical application of that knowledge (Gorbacheva et al., 2016).

2.1 Required competences of data professionals

There have been recent attempts to clarify competence requirements for big data in general (Debortoli et al., 2014; Dubey and Gunasekaran, 2015) and for data analytics in supply chain management (Schoenherr and Speier-Pero, 2015), as well as for upcoming data-driven occupations, such as those of mobile analysts (Brauer and Wimmer, 2016) and data scientists (Davenport and Patil, 2012; Schumann et al., 2016). Beyond these attempts, there is no scientific work regarding individual competence requirements, although a thorough understanding seems to be important for developing promising curricula of university programmes for educating data professionals.

According to Davenport and Patil (2012), a data scientist is ‘a high-ranking professional with the training and curiosity to make discoveries in the world of big data’ (p. 72). The occupation name itself provides a first hint of what a data scientist is. Data indicates the major working object that is a mass of unstructured data that needs to be brought into an analysable format. The scientist refers to a specific mind-set characterised by curiosity and the desire to create new things beyond ‘simple’ consultancy (Davenport and Patil, 2012). There is general agreement on the fact that data professionals need to have multidisciplinary competences (Zawadzki, 2014; Schoenherr and Speier-Pero, 2015). Based on a literature review of 17 publications, Schumann et al. (2016) present an overview of data scientist competences clustered by the competence categories labelled  
professional  (e.g., statistics or skills for the selection, pre-processing, analysis, and interpretation of data),  
social  (e.g., teamwork and communication), and  
personal  (e.g., curiosity or creativity). Zawadzki (2014) distinguishes between the four main categories of  
math & statistics, programming & database, domain knowledge & soft skills, and communication & visualization  and provides examples for all of them. Schoenherr and Speier-Pero (2015)
mention that a data scientist must possess competences of enterprise business processes and decision-making, data management, and analytical and modelling tools. Generally, these results underpin the image of an ‘all-rounder’. To overcome the issue of finding one person possessing all required competences, many scholars emphasise the establishment of a data science team (Schumann et al., 2016; Waller and Fawcett, 2013; Zawadzki, 2014). With respect to its novelty, most of the existing publications on the topic of data professionals’ competences have followed rather exploratory approaches. For instance, Debortoli et al. (2014) state that specific big data job profiles are often only anecdotally described.

One rigorous scientific approach to overcome the shortcomings concerning insights into data professionals’ competences would be the development of a typology similar to the one of Müller et al. (2014) who based their typology of business process management professionals on job advertisements (job ads). Generally, typologies can be defined as a classification into structural types of the phenomenon under study (Croft, 2003). According to Doty and Glick (1994) who argue for a strict distinction between typology, classification, and taxonomy, typology ‘refers to conceptually derived interrelated sets of ideal types’ (p. 232). Thus, by developing a typology, multiple ideal types are identified as types that are expected to be effective or successful. Typologies state relationships between independent variables (ideal types) and dependent variables (e.g., success, performance), thereby going beyond pure description (Müller et al., 2014). Our approach is to inductively develop a typology of data professionals (see also section 3.1). Following the guidelines for developing typologies proposed by Doty and Glick (1994), it is necessary that each ideal type is described using the same set of dimensions. For this reason, Todd et al.’s (1995) approach serves as the basis for this present study (see section 2.3).

2.2 Design and assessment of data-related curricula

The question of whether academic education meets presumed IS and IT industry needs in general has been investigated for some time (Lee et al., 1995; Noll and Wilkins, 2002; Targett, 1991). Typically, these studies have considered IS jobs without any further distinction. With regard to the increased role of specific IS components such as big data, cloud computing, or cyber-physical systems, a more detailed view is required. Scholars have started to address the topic of data-related curricula in recent years. This has been caused by a growing demand for data professionals and increased offerings of data-related study programmes.

Corresponding publications often deal with ‘calls to action’, meaning that data-related curricula need to be established (Dubey and Gunasekaran, 2015; Wixom et al., 2014; Wixom et al., 2011), thereby reflecting the novelty of this research stream. Some studies deal with design proposals for and first assessments of data-related curricula (Gupta et al., 2015; Mitri and Palocsay, 2015).

Turel and Kapoor (2016) have analysed 124 business schools in the US regarding their business analytics offerings. Specifically, they consider single courses offered and assign each to one of four predefined categories: database, data analysis, business analytics, or data warehousing. Furthermore, they distinguish between ranked and unranked business schools and provide several descriptive evaluations. Their core finding is that the business schools they analysed largely did not reach acceptable levels of business analytics maturity, meaning that presumed industry needs were not met. However, what all the above approaches have in common is that they assume certain competences that a data professional should possess and also derive curricula recommendations, but they do not determine
actual demanded competences. One exception is Schoenherr and Speier-Pero (2015) who focus on data professionals working in the supply chain management sector. Among others, they investigate required competences (via online surveys and expert interviews) before analysing how US universities train their data professionals. They consider entire master’s programmes and assign courses (only mandatory ones) to one of three previously determined categories (enterprise business processes and decision making, analytical and modelling tools, data management).

However, Schoenherr and Speier-Pero (2015) do not provide a comparison of demanded and imparted competences. What they provide instead is the evolution of the curricula of a master’s programme called ‘Predictive Analytics’ at the Michigan State University (US).

### 2.3 The IS competence framework of Todd et al.

Todd et al. (1995) deal with the evolution of IS job competences that are analysed based on job advertisements published in US newspapers. They classify competences listed in job ads into three main categories: technical, business, and system. Sub-categories and their descriptions are displayed in Table 1.

<table>
<thead>
<tr>
<th>Category</th>
<th>Sub-Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technical</td>
<td>Hardware</td>
<td>Servers and personal computers. Other devices such as storage devices, controllers, printers, and other peripherals, plus networks.</td>
</tr>
<tr>
<td></td>
<td>Software</td>
<td>Application systems, operating systems, packaged products (e.g., databases), networking software, and programming languages.</td>
</tr>
<tr>
<td>Business</td>
<td>Domain</td>
<td>Functional expertise (e.g., finance, marketing) and industry expertise (e.g., retail, mining).</td>
</tr>
<tr>
<td></td>
<td>Management</td>
<td>General management skills including leadership, project management, planning, controlling, training, and organisation.</td>
</tr>
<tr>
<td></td>
<td>Social</td>
<td>Interpersonal skills, communication skills, personal motivation, and ability to work independently.</td>
</tr>
<tr>
<td>System</td>
<td>Problem-Solving</td>
<td>Creative solutions, quantitative skills, analytical modelling, logical capabilities, deductive/inductive reasoning, innovation.</td>
</tr>
<tr>
<td></td>
<td>Development</td>
<td>Knowledge of systems development methodologies, systems approach, implementation issues, operations and maintenance issues, general development phases, documentation, and analysis/design tools and techniques.</td>
</tr>
</tbody>
</table>

*Table 1: Classification of IS competences (Todd et al. 1995, p. 6).*

Although there are other sets of IS competence dimensions (e.g., Litecky et al. (2009) who propose a classification for ‘MIS skills’), we decided to select Todd et al.’s (1995) approach for the following reasons: First, although their classification scheme is more than 20 years old, it offers a relatively broad framework covering technical as well as business aspects that are also applicable to today’s working environment. Second, they used the classification scheme in a similar context to ours, which is job ads analysis. Third, this approach has recently been applied successfully by other scholars fo-
cussing on specific occupations such as business process management professionals (Müller et al., 2014). Fourth, besides the typology context, the framework is also applicable to providing categories and sub-categories for the deductive content analysis of master’s programmes (see section 3.2).

3 Research Approach

Figure 1 provides an overview of the general research approach for this study. Eventually, two independent analyses, one dealing with required competences and the other with imparted competences, are conducted in the first step. These analyses differ regarding data and methods. The linking element is the IS competence framework of Todd et al. (1995) that allows comparability (see also sections 2.3 and 4) in the second step. With respect to the novelty of applying a topic model algorithm in the IS field (Müller et al., 2016), related sections (section 3.1) are presented in more detail than the already established deductive content analysis (section 3.2).

Figure 1: General research approach.

3.1 Job advertisement analysis

To gain a deeper understanding of data-related occupations regarding required competences and associated tasks, we have chosen an explorative approach: an analysis of data-related job ads. A job ad is comprised of the obligatory qualifications and competences required to fulfil the tasks and activities that arise during work and business processes (Baron et al., 2009) and has been considered data by other IS scholars for identifying competences (e.g., Todd et al. (1995); Müller et al. (2014); Debortoli et al. (2014); Ha (2016)). Furthermore, because of the detailed information on employer-demanded competences, job ad data are seen as a valuable source for guiding curriculum developers (Carnevale et al., 2014). Today, the majority of job ads are accessible via online job search platforms and a large collection of documents (that differ regarding their structure) can be easily downloaded. In fact, a manual content analysis of such a large data set (depending on search parameters, hundreds of job ads might be available) seems to be impractical. Therefore, we follow recently published recommendations (Debortoli et al., 2016; Müller et al., 2016) and apply a probabilistic topic model approach, namely the Latent Dirichlet Allocation (LDA) introduced by Blei et al. (2003).
3.1.1 Data collection and data preparation

The main aspects of collecting job ads from the Internet are displayed in Table 2. Our combination of search terms resulted from different search term tests. The chosen terms ensure coverage of all relevant occupations such as data scientist, data analyst, business intelligence analyst, or digital analyst. Five hundred job ads were downloaded with a download manager application. The following steps describe data preparation and cleansing. First, we reviewed the sample regarding job ads that were out of scope (e.g., we identified a few job ads that dealt with chemical or biological data analysis) and removed those. Subsequently, we applied a duplication finder tool for identifying those job ads with a similarity of more than 80% to one or even more job ads in our sample. All hits were checked manually and removed if necessary. These steps led to a total sample of 459 different job ads.

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Our choice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Online job search platform</td>
<td><a href="http://www.monster.co.uk/">http://www.monster.co.uk/</a></td>
</tr>
<tr>
<td>Selected region</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>Download date</td>
<td>24 October 2016</td>
</tr>
<tr>
<td>Search terms</td>
<td>Data Analytics + Data Analyst</td>
</tr>
<tr>
<td>Job ads downloaded</td>
<td>500 (sorted by relevance)</td>
</tr>
</tbody>
</table>

Table 2: Overview of job ads collection.

In the next step we deleted those parts of the job ads that were not relevant for our purposes (e.g., company information and contact details or salary information). Furthermore, we applied common cleansing techniques for text corpora as described in Debortoli et al. (2016) such as eliminating standard stop words (removing common or uninformative words), lemmatization (reducing a word into its dictionary form; e.g., plural to singular for nouns, verbs to simple present tense), and stemming (reducing a word to its stem). Subsequently, the text corpus was analysed by using descriptive statistics that revealed 53,457 words and 1,459 unique words. Following the approach of Müller et al. (2014), two researchers independently checked the list of unique words for identifying those words not relevant for our purposes (so-called custom stop words; e.g., caused by our search terms, the word ‘data’ has a very high frequency but does not add particular value). They agreed in 81% of the cases. The remaining cases were clarified during a discussion leading to a total set of 226 words that were disregarded for further investigations.

3.1.2 Data analysis

While a detailed examination of the underlying mathematics is beyond the scope of this study, the main ideas and assumptions of LDA are briefly presented. Generally, the aim of LDA is to discover the main themes (‘topics’) that pervade a large and otherwise unstructured collection of documents (Blei, 2012). Figure 2 illustrates the relation between topics, (unique) words, and documents (here: single job ads). According to Debortoli et al. (2016), ‘a probability distribution over a fixed set of topics defines each document, and, in turn, a probability distribution over a confined vocabulary of words defines each topic’ (p. 114).
We used the LDA-based text mining application MineMyText (http://www.minemytext.com/) for analysis. As LDA requires a given number of topics, we conducted several computations with different numbers of topics. A small number led to very broad and generic topics, and a high number led to very specific or even niche topics. Similar to Müller et al. (2014) who have chosen a number of seven for their business process management typology, we decided to have five topics. Table 3 displays the results of the job ad analysis with a given number of five topics that we labelled as ideal types of data professionals according to our research goal. As usual for topic model results, we also report the terms with a high proportion per topic as well as titles of job ads with a high proportion per topic. Based on their interpretation, ideal type names are determined.

<table>
<thead>
<tr>
<th>#</th>
<th>Ideal Type Name</th>
<th>Terms with High Proportion (Sorted by Proportion Value, Stemmed)</th>
<th>Titles of Job Ads with High Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Internal Processes Specialist</td>
<td>Report, excel, support, market, inform, team, perform, communic, ensur, plan, develop, forecast, improve, financ, product</td>
<td>Business Planning Analyst, Business Information Analyst, Operational Analyst</td>
</tr>
<tr>
<td>2</td>
<td>Online Marketing Specialist</td>
<td>Googl, market, web, report, understand, test, tag, perform, tool, drive, adob, develop, custom, websit, social</td>
<td>Digital Marketing Data Analyst, Social Analyst, Online Optimisation Analyst, Social Media Analyst</td>
</tr>
<tr>
<td>3</td>
<td>Customer Insights Specialist</td>
<td>Custom, team, sa, model, market, stakehold, sql, deliv, statist, lead, project, excel, strategi, recommend, communic</td>
<td>SQL Insight Analyst, Lead Analyst, Customer Insight Analyst</td>
</tr>
<tr>
<td>4</td>
<td>Functional Support Specialist</td>
<td>Support, system, team, process, service, ensur, issu, project, develop, inform, chang, understand, oper, custom, implement</td>
<td>Financial and Business Application Support Analyst, Information Security Analyst, HR Systems and Process Analyst</td>
</tr>
<tr>
<td>5</td>
<td>Programming Specialist</td>
<td>Develop, sql, report, technic, solut, excel, team, communic, design, database, support, stakehold, deliv, project, problem</td>
<td>Data Solution Architect, Big Data Developer, MI Data Developer, Software Engineer</td>
</tr>
</tbody>
</table>

Table 3: Results of job ad analysis.
### 3.1.3 Linking with the IS competence framework of Todd et al. (1995)

Given the results of LDA analysis, the link to the IS competence framework of Todd et al. (1995) can be established, thereby ensuring the postulation of Doty and Glick (1994) that each ideal type is described using the same set of dimensions. In detail, the 30 terms with the highest proportions are assigned to one of the sub-categories (Table 4). Two researchers conducted this coding procedure. They agreed in 77% of the cases; the remaining cases were clarified during a discussion.

<table>
<thead>
<tr>
<th>Category</th>
<th>Sub-Category</th>
<th>Assigned Terms</th>
<th>Assigned Terms</th>
<th>Assigned Terms</th>
<th>Assigned Terms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td><strong>Internal Processes Specialist</strong></td>
<td><strong>Online Marketing</strong></td>
<td><strong>Customer Insights Specialist</strong></td>
<td><strong>Functional Support Specialist</strong></td>
</tr>
<tr>
<td>Technical</td>
<td>Hardware</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Software</td>
<td>excel, googl, web, tool, adob, web,</td>
<td>excel, sql, excel</td>
<td>excel, applic</td>
<td>sql, excel, databases, tool, microsoft, bi, platform</td>
</tr>
<tr>
<td></td>
<td>Business</td>
<td>market, financ, intern, product,</td>
<td>market, custom, onln, user</td>
<td>custom, market, segment, environ, product</td>
<td>service, custom, hr, secur, risk commerce</td>
</tr>
<tr>
<td></td>
<td>Management</td>
<td>report, support, inform, plan, forecast, time, decis, trend</td>
<td>report, campaign, strategi</td>
<td>lead, project, strategi, campaign, action</td>
<td>support, project, inform, chang, report, respons, document report, support, project, respons</td>
</tr>
<tr>
<td></td>
<td>Social</td>
<td>team, ensur, communic, level, present</td>
<td>drive, social, channel, present</td>
<td>team, stakehold, commun, drive, level, behaviour</td>
<td>team, ensur, communic, group team, communic, stakehold</td>
</tr>
<tr>
<td></td>
<td>System</td>
<td>improv, research, effect, accur, understand</td>
<td>understand, optimis, improv, convers, recommend</td>
<td>deliv, statist, recommend, understand, numer, opportun,</td>
<td>issu, understand, effect, problem, appropri solut, deliv, problem, complex, source, understan</td>
</tr>
<tr>
<td></td>
<td>Development</td>
<td>perform, develop, oper, identifi, process</td>
<td>test, tag, perform, develop, implement, identifi, track</td>
<td>model, develop, build, perform, advanc</td>
<td>system, process, develop, oper, implement, perform, qualiti, identifi develop, technic, design, model, technolog, implement, user, system, process, build</td>
</tr>
</tbody>
</table>

*Table 4: Linking ideal types to IS competence framework of Todd et al. (1995).*

### 3.2 Master’s programme analysis

One objective of master’s programmes is to build or increase employability. In other words, master’s programmes serve the job market by providing well-educated people. The market is evolving rapidly for master’s programmes that focus on the topic of data. For instance, while Davenport and Patil (2012) note that there are no university programmes offering degrees in data science, we presently observe an increase in related programme offerings. What often remains uncharted is the assessment of how specific curricula meet presumed industry needs. The IS field offers some initial approaches to answering this question (Schoenherr and Speier-Pero, 2015; Turel and Kapoor, 2016). Our analysis is designed in a similar fashion.
3.2.1 Data collection and data preparation

Following Schoenherr and Speier-Pero (2015), we focussed on Master of Science (MSc) programmes. This is in line with the context of our study, which is labour-relevant competences: compared to undergraduate programmes (i.e., bachelor’s), master’s graduates are ‘closer’ to the labour market. Furthermore, for deriving representative findings, a certain number of data-related programmes is required; this is not the case for other potentially relevant programmes such as PhD programmes. The search for data-related master’s programmes has been conducted through a course search engine for higher education (http://www.mastersportal.eu/). This database lists over 50,000 master’s degrees from over 3,300 universities, colleges, and graduate schools all over the world. While applying the proposed search terms Data Science & Big Data and Business Intelligence & Analytics and focussing on the UK, 43 relevant master’s programmes were identified (as of November 2016). These programmes were manually reviewed, particularly regarding the available information on the course level that is needed for further analysis (see 3.2.2). Furthermore, those programmes with too narrow of a purpose (e.g., MSc in Data Telecommunications and Networks) were excluded. Finally, we selected 12 master’s programmes for detailed analysis.

3.2.2 Data analysis

Obviously, the sample size of 12 master’s programmes requires a different analysis approach than the one for hundreds of job ads (see section 3.1). Thus, a qualitative content analysis (Mayring, 2014) was used as an analysis method. Each programme was systematically examined on the course level while applying a deductive category assignment. This deductive approach is in line with existing curricula assessment studies (Schoenherr and Speier-Pero, 2015; Turel and Kapoor, 2016). One of the core steps of deductive category assignment is the definition of the category system that contains both categories and sub-categories (Mayring, 2014). For this purpose, and similar to the description of the data professional ideal types (see section 3.1.3), the IS competences scheme of Todd et al. (1995) serves as a foundation. It must be noted that we analysed both compulsory as well as optional courses (except the master’s thesis modules) to represent the overall offerings per programme. In addition, we assumed that the curricula for such a young discipline have a dynamic character, meaning that the course types (compulsory vs. optional) are often subject to change. As we furthermore assumed that one course covers several sub-categories or even categories, we decided to have three ‘votes’ per course to be assigned to the sub-categories. For instance, in case only one sub-category was suitable, all three votes for the course were assigned to this sub-category. In most cases, the votes were spread between the sub-categories. Also, this coding procedure was conducted by two researchers independently. They agreed in 86% of the cases and clarified the remaining cases during a discussion.

4 Merged Results and Discussion

Figure 3 summarizes our findings and illustrates both the five extracted ideal types of data professionals and their required sets of competences on the left side as well as the 12 analysed master’s programmes including those sets of competences that are offered on the right side. According to the IS competences set of Todd et al. (1995), the blue parts belong to the technical category, the red parts to business, and the green parts to system. This discussion is guided by our three research questions.
Figure 3: Comparison of demanded and imparted data-related competences.

RQ1: What are required competences of data professionals in the UK? Considering the five ideal types of data professionals, the graphical analysis uncovers different competence requirements. Generally, all ideal types require substantial business competences. Furthermore, all of them require profound development and problem-solving competences while hardware competences are not required at all. Specifically, the internal processes specialist should possess a strong management competence; this intuitively makes sense due to related project management and communication tasks. This ideal type needs only few software competences. Online marketing specialists require less business competences than internal processes specialists, but should possess substantial software competences (e.g., Google analytics and Adobe analytics). Customer insights specialists require solid competences for all three business competences but need fewer software competences. Similar to internal processes specialists, functional support specialists require strong management competences. The programming specialist must possess very strong system competences and show less business competence requirements compared to the other ideal types. Generally, the extracted ideal types cannot be assigned to one category (such as pure technician) but show the image of an ‘all-rounder’, thereby confirming existing literature (Schumann et al., 2016; Zawadzki, 2014).

RQ2: What competences are imparted by data-related master’s programmes in the UK? The results of the master’s programme (MP) analysis is illustrated on the right side of Figure 3. It must be noted that these results reflect offerings of single master’s programmes, which is not necessarily the same as the actual competences of the graduates. This is because we analysed all courses in a master’s programme, be it compulsory or optional. Furthermore, it is noteworthy that we analysed intended learning outcomes stated in course descriptions that are not equal to actual imparted competences. Generally, the most prominent competences offered by master’s programmes are software, problem-solving, and development. The graphical analysis illustrates that all programmes contain a relatively small share of business competences especially compared to problem-solving and development competences. One important question is whether this is on purpose or not. If it were on purpose, this might be caused by the ‘specialization’ character of a master’s programme. Business competences could be expected to be
taught in undergraduate programmes. However, persons in charge of the specific curricula would have to be interviewed to confirm our conclusion.

**RQ3: How do data-related master’s programmes match competence requirements for data professionals in the UK?** Our comparison shows that the competences imparted by data-related master’s programmes do not fit all the ideal types. This is generally in line with the findings of Turel and Kapoor (2016). Current master’s programmes focus on system (development, problem-solving) competences which might be interpreted as the foundation for a creative, critical, and analytical way of working. While the share of technical competences seems to fit well, particularly for online marketing and programming specialists, business competences tend to be generally underrepresented. This is not surprising for domain competences as we have excluded those master’s programmes with too narrow of a focus (see section 3.2.1). Furthermore, this seems to be a classical field that is learned on the job. Yet, management (e.g., project management, reporting) and social (e.g., team, communication) competences could easily be stressed in curricula. As discussed before, there seem to be reasons not to focus on business competences. In any case, due to great demand, universities might adjust their curricula accordingly. Another more employability-oriented way of improving fit to market demand could be the focus on single ideal types, such as a specific programme for customer insights specialists, etc. Current programmes with a specialization mainly focus on industries instead. Such a new ideal-types-oriented specialization approach would, however, require substantial changes to existing curricula.

### 5 Conclusion and Future Work

The main theoretical contribution of this paper is the development of an empirically derived typology of data professionals based on job ads. Also, by making use of a topic model algorithm, IS research is advanced from a method-related perspective (Müller et al., 2016). Additionally, some practical contributions can be presented. Considering both aspects in one integrative approach allows a discussion of whether existing curricula meet industry demands or not based on empirical findings. By doing so, educational institutions are enabled to critically assess their offerings according to the graduate’s employability aspect, and this might shape their offerings correspondingly.

One of the major limitations of the topic model approach we have applied is the lack of quality measurements (Debortoli et al., 2016; Lau et al., 2014; Müller et al., 2016). Although scholars such as Boyd-Graber et al. (2014) and Lau et al. (2014) have addressed this issue, there are no ‘significance tests’ or other established evaluation procedures at this time. However, because of the increased meaning of innovative big data analytics such as topic models for the IS field (Müller et al., 2016), we believe that these issues will be solved in the near future. Besides these method-related challenges, our data show some limitations. We have analysed a snapshot (of both job ads and curricula) that enables us to draw conclusions about the current state while historical developments or future requirements cannot be investigated. Furthermore, in our study, there is no information about how single job advertisements have been created (e.g., if only standard phrases are used, etc.). Also, as some employers still continue to use traditional methods of advertisements such as newspapers and career fairs or try to deploy employees’ social networks, online job ads do not represent a complete picture of demand for a specific occupation (Carnevale et al., 2014). These limitations lead to our call to include HR and domain experts who could shed light on these questions. In addition, we must notice that formal education such as study programme offerings are only one component to developing competences of data
professionals. There are other important means such as informal education that is assumed to have a direct linkage to hard as well as soft competences (Dubey and Gunasekaran, 2015). Finally, when considering the shortage of professionals for job markets, study programmes alone cannot solve the problem. Instead, further actions such as occupational re-training need to be established.

Future research might benefit from the presented analysis approach as it can easily be applied to different contexts (e.g., different countries or occupations) without major adjustments. For instance, cross-country analyses could be conducted. Also, while our study considers both job ads and master’s programmes on a rather aggregated level (i.e., ‘data-related’), it might be valuable to narrow down the focus to a specific industry or function, for example. Another interesting aspect could be to research the evolution of a specific occupation, comparable to the approach of Todd et al. (1995) for IS jobs framed in a longitudinal study. In an LDA context, this would mean that the ordering of the document is considered relevant, which is not the case for standard LDA applied in this study. However, dynamic topic models have been developed to tackle such issues (Blei, 2012). Finally, the empirically derived typology of data professionals might be subject to external validation (e.g., by interviewing domain experts).

References


