



Paper to be presented at  
DRUID15, Rome, June 15-17, 2015  
(Coorganized with LUISS)

## **Using websites to reveal knowledge types: An innovation in research methods**

**Katia Oliveira Pina**

The University of Manchester - Manchester Business School  
Manchester Institute of Innovation Research  
katia.pina@mbs.ac.uk

**Bruce S. Tether**

University of Manchester  
Manchester Business School  
bruce.tether@mbs.ac.uk

### **Abstract**

Knowledge has long held a central place in innovation studies, and in recent years Asheim and colleagues (2005) have developed a conceptualisation based on three different 'types of knowledge', that goes beyond the established distinction between 'tacit' and 'codified' knowledge. Very few studies have however either operationalised this conceptualisation, or asked whether other 'types of knowledge' exist and can be identified. This paper does both, applying innovative methods to extract information from company websites. In particular, and by analysing website content information from 450 UK based professional business service firms, this paper makes the four contributions. First, it identifies a hitherto unmeasured 'type of knowledge': compliance knowledge; Second, it develops a method for identifying types of knowledge; Third, it applies this method to textual content extracted from company websites; Fourth, it tests this evidence, both against the characteristics of the firms included in the analysis, and against survey evidence. We conclude that the method enables valuable and insightful information that complements existing categories (e.g., sectors) concerning a key competitive aspect of these firms. Furthermore, this could motivate further research on conceiving, identifying and distinguishing different 'types of knowledge'.

# Using websites to reveal knowledge types (among Professional Service Firms): An innovation in research methods

Names of authors and their affiliations removed.

## ABSTRACT

Knowledge has long held a central place in innovation studies, and in recent years Asheim and colleagues (2005) have developed a conceptualisation based on three different ‘types of knowledge’, that goes beyond the established distinction between ‘tacit’ and ‘codified’ knowledge. Very few studies have however either operationalised this conceptualisation, or asked whether other ‘types of knowledge’ exist and can be identified. This paper does both, applying innovative methods to extract information from company websites. In particular, and by analysing website content information from 450 UK based professional business service firms, this paper makes the four contributions. First, it identifies a hitherto unmeasured ‘type of knowledge’: compliance knowledge; Second, it develops a method for identifying types of knowledge; Third, it applies this method to textual content extracted from company websites; Fourth, it tests this evidence, both against the characteristics of the firms included in the analysis, and against survey evidence. We conclude that the method enables valuable and insightful information that complements existing categories (e.g., sectors) concerning a key competitive aspect of these firms. Furthermore, this could motivate further research on conceiving, identifying and distinguishing different ‘types of knowledge’.

**Keywords:** knowledge types, content analysis, websites, mixed methods.

**Suggested tracks:** C. Innovation strategy and organizational behaviour

D. The firm, organization and markets

Body of Text Word Count: 8,236

Total word count including notes, tables, appendices, references, etc. = 10,994

## **1. INTRODUCTION**

Knowledge has a central role in the literature on innovation and innovation systems (e.g., Pavitt, 1984; Lundvall, 1993; Edquist, 1997; Malerba, 2002, 2005; Asheim & Gertler, 2005), and in recent years innovation scholars have been conceptualising different ‘types of knowledge’. While much of this work is conceptual, and applied at high levels of aggregation, such as at the sector or regional levels, more recently attempts have been made to apply the concept of ‘types of knowledge’ at the firm level, particularly with reference to knowledge intensive, or professional service firms (Pina & Tether, 2015).

This paper aims to advance that stream of work further, by attempting to measure the presence, strength or prominence of different ‘types of knowledge’, and does this by extracting information from company web sites. We do this for around 450 UK based professional business service firms. The paper makes four principal contributions. First, it identifies a hitherto unmeasured ‘type of knowledge’: compliance knowledge; Second, it develops a methodology for identifying ‘types of knowledge’ from text based evidence; Third, it applies this methodology to information drawn from company websites; Fourth, it tests this evidence, both against the characteristics of the firms included in the analysis, and against survey evidence.

The paper is structured as follows. Section 2 provides an overview of the literature on ‘knowledge types’, especially as developed by Bjorn Asheim and colleagues. Section 3 then discusses websites as a source of information about firms, and their characteristics. A fundamental assumption of this paper is that websites can reveal at least something about the ‘knowledge types’ being utilised by firms. Section 4 discusses our methods and some initial findings, while Section 5 reports the analysis of the findings which are modelled against company characteristics. Finally, section 6 provides a brief discussion of the findings and concludes the paper.

## **2. CONCEPTUAL BACKGROUND**

While several taxonomies of knowledge have been proposed (Kakabadse et al., 2003), Asheim and colleagues (2005) have developed an epistemological classification that goes beyond the traditional distinction between tacit and codified knowledge (Polany, 1967).

Indeed, the three ‘types of knowledge’ identified by Asheim and colleagues not only represent different combinations of tacit and codified knowledge but also reflect different rationales for and approaches to knowledge creation (Martin, 2013). They are considered to be ‘ideal types’ (Asheim and Hansen, 2009, p.431), and several ‘types of knowledge’ can co-exist at different levels of analysis, including regions, industries and firms. This said, particularly at the micro level, some firms can be expected to specialise in a specific ‘types of knowledge’ (Tether et al., 2012), as these relate to organisational structures and identities (Kogut and Zander, 1996). The concept has been applied empirically, mainly to differentiate regions and industries in relation to their innovation behaviours (e.g. Asheim & Coenen, 2005; Moodysson et al., 2008; Coenen & Moodysson, 2009), but very few studies have sought to measure knowledge types; an interesting exception being Martin (2012), which used occupation data in association with location quotients to operationalise the concept.

We now briefly review the main characteristic of the three types of knowledge identified by the ‘SAS model’: Synthetic, Analytical and Symbolic.

‘Synthetic’ knowledge is primarily developed to address or solve a specific problem. Rather than consisting of abstract concepts and systematic methods, synthetic knowledge is essentially practical, solution oriented, and ad hoc (Asheim et al., 2005). What matters most is that the solution works; it does not have to be the optimal solution; nor does it matter (in the first instance) whether or not the solution used can be replicated. This said, synthetic knowledge draws heavily on experience, so applications found to be suitable in one context are likely to be tried in other, similar contexts.

‘Analytical’ knowledge is considered to be based on deductive logic, and requires skills related to the ability to work with abstract concepts and engage in empirical testing. It is largely developed using recognised, predefined or legitimated ‘scientific’ methods that are framed by systematic and organised structures and relationships. R&D activities are, for example, typically analytical (Asheim et al., 2005).

Asheim and colleagues distinction between analytical and synthetic knowledge shares many parallels with Gibbons and colleagues’ (1994) distinction between Mode 1 (scientific) and Mode 2 (practical problem solving) knowledge. The main ‘innovation’ in Asheim et al’s conceptualisation is the addition of ‘symbolic knowledge’, which concerns the creation, development and manipulation of signs, symbols and other forms of expression, including

images, narratives and sounds. This ‘type of knowledge’ is especially relevant to semiotically rich activities, such as the media, fashion, design and advertising (Asheim et al., 2005). For this ‘type of knowledge’, there are two primary requirements: 1. The ability to interpret, manipulate and create symbolic forms of expression, and 2. The ability to persuade others of the value of new signs and symbols. Table 1 provides a summary of the distinctions between these three ‘types of knowledge’.

-----  
Insert Table 1 about here  
-----

This conceptual distinction between knowledge types is interesting, but raises questions. One is that there may be other ‘types of knowledge’, in addition to these three types. For example, what about knowledge about laws, regulations, and compliance with ‘proper procedures’? Such knowledge does not fit comfortably into any of the SAS types. Does compliance knowledge therefore constitute a separate, as yet unmeasured ‘type’ of knowledge? Another question is what is the relationship between these different types of knowledge? For example, some may be closer together, and others more distant. In other words, it may be easier for a firm primarily engaged in knowledge type A to diversify into activities based on knowledge type B as opposed to those based on knowledge type C. An understanding of this would aid understanding of the direction of diversification that firms typically take.

Separate from the conceptual challenge is the practical challenge of empirically identifying and measuring ‘types of knowledge’. As mentioned, very few studies have directly measured these. Two exceptions are Martin (2012) and Tether et al., (2012), both of which assume that different occupations relate to particular ‘types of knowledge’.

To develop research in this area, we sought to identify and measure four ‘knowledge types’ by drawing from textual information provided by firms. Specifically, we drew on the textual information that firms make public on their websites. In an attempt to validate our findings, we later cross checked the results against other information available to us, including information about firm size, sector of activity and the opinion of a survey respondent on the importance of different types of knowledge to the activities of the firm.

### **3. WEBSITES AS SOURCES OF INFORMATION ABOUT FIRMS**

The methods we outline below could in principle be applied to any rich textual accounts that firms provide about themselves and their activities, for example annual reviews (which often accompany company accounts), or documents prepared by their parties (company profiles, company histories, case study materials). For this analysis, we utilise the information that firms make public on their websites. This is advantageous, because beyond a certain minimal size, almost all firms in advanced economies now have a website, which they use to engage with customers or clients, and other stakeholders. While websites serve multiple functions, they allow the firm to communicate with different audiences through a single channel (Hwanga et al., 2003). Within the law, firms are free to present whatever ‘face to the world’ they wish to present, which requires decisions about the amount and type of information, and about presentation (e.g., highly graphical, highly textual, etc.). Because website space is essentially infinite, and website hosting costs are low, company’s websites can be of any size. Key decisions are likely to include how dynamic website should be (i.e., how frequently its content is updated) and the extent to which the website is designed to attract attention (and from whom), particularly through search engine optimization techniques.

There are of course scientific literatures on both corporate communications and more specifically on websites and their use. In relation to the corporate communications literature, emphasis is often placed on communications being a source of ‘controllable’ information, often managed by marketing and public relations within or on behalf of the firm. Such communications reflect firms’ “communicated identity” (Balmer and Greyser, 2002). This corresponds to ‘what organisations say they are’ (Bronn and Martisen, 2006), or how they want to be perceived, rather than a full account of their actual identity (Balmer and Greyser, 2002) or “substance” (Alvesson, 1990). While firms are unlikely to wholly misrepresent themselves, they are likely to accentuate those aspects that they consider to be attractive. Firms will aim to show they are doing, or are able to do, a good job for clients and other key stakeholders. The ‘substance’ and the ‘communicated identity’ are therefore different, but ultimately closely connected (Alvesson, 1990). In practical terms, this means that we should not expect the websites to provide a full and unbiased account of the activities of the firm, but rather to accentuate those aspects (in our case, ‘knowledge types’, that are of greatest significance).

Meanwhile, the specific literature on corporate websites considers these to be public data sources that provide information regarding the firm's main activities, strategies and identity (Gioia and Thomas, 1996; Scott and Lane, 2000). Several researchers have drawn textual and visual content from websites to infer information about firms' corporate identities (e.g., Knox and Bickerton, 2003; Baak and Singh, 2007; Cornelius et al, 2007; Haniffa and Hudaib, 2007; Rodrigues and Child, 2008). Researchers have emphasised the advantages of examining websites, including that it is an 'unobtrusive' technique, and that the data – and especially text – can be easily captured and 'read' by computer programmes. A disadvantage is loss of control concerning the information available, which varies enormously in amount, content and quality. Content analysis can however provide a systematic way of identifying distinctive content features (King, 2004; Neuendorf, 2007) and has been used as a valid way of identifying firm's *raison d'être* (Stemler, 2001).

Generally, larger firms have more sophisticated, information rich websites, partly because of the greater resources available, but also because they are more likely to be communicating with multiple audiences, and managing relationships with multiple stakeholders (Hwang et al., 2003). But more important than firm size, is the nature of the audience. Firms dealing with consumers tend to have more sophisticated, and more frequently updated, websites, as this has become an important channel for mass communications. In our case the firms we were interested in are professional service firms, providing services to other businesses, rather than the general public. Although their websites are likely to be an important 'front door' to the business (for prospective employees, as well as clients), websites were not the primary channel through which these firms interacted with its clients; indeed, survey evidence shows they still have a very high reliance on face to face interactions, especially when trying to win new business.

During our research, and indeed partially as a motivation to subsequently explore this more systematically, we noticed some striking differences between the website content and styles of companies in different sectors. Engineering consulting firms, for example, were typically presented in a 'serious', authoritative way, with sober colours, headings, and text. Quite frequently the names of key personnel were not provided, and when they were they were almost always presented in a formal way, e.g., Mr John Smith, or Dr Gordon Brown. By contrast, firms in more 'creative' or 'expressive' sectors tended to present themselves differently – with lots of images, strong use of colour, and in a highly informal, or fun way.

For example, the surnames of key personnel were frequently not given, just first names. Quite often the office pet was presented as if it were another member of the workforce. The work of the firm was also typically presented in a manner that closely identified the work with the people directly involved.

Our challenge was to extract information in a systematic, ‘scientific’ way, with a view to identifying distinct ‘types of knowledge’ which cannot be observed directly, but which we assume to be latent variables.

#### **4. METHODS AND INITIAL FINDINGS**

Partly inspired by the casually observed differences mentioned above, we sought to systematically extract textual information provided on companies’ websites with a view to identifying their engagement with different, not directly observable, types of knowledge. This involved several challenges which we will outline in this section. Later, we cross check our findings against the characteristics of the firms and against answers to a survey question concerning the types of knowledge used by the firms. Figure 1 provides a graphical representation of our research process.

-----  
Insert Figure 1 about here  
-----

Our starting was the 541 firms that provided responses to the aforementioned survey of professional and business service firms. These were architecture practices; creative advertising agencies and design consultancies; engineering consultancies; and market research firms. Importantly, this set of firms could be expected to exhibit variety in the ‘types(s) of knowledge’ at the core of their activities. For example, we anticipated that market research firms would depend most heavily on ‘analytical knowledge’, while advertising/design firms would rely heavily on ‘symbolic knowledge’, as would architecture practices, albeit possibly to a lesser degree. Engineering consultancies could be expected to show a strong orientation to ‘synthetic’ and ‘compliance knowledge’, and perhaps ‘analytical knowledge’ to a lesser degree.

First, we identified websites for the firms, finding these for all but four.

The next challenge was extracting information from these websites, particularly in light of the huge diversity in website sizes and quality. In order to overcome this, some researchers have opted to focus on specific pages, such as the home page, but this approach can be arbitrary (which pages should be included, and which excluded?), and also greatly reduces the amount of information analysed. In this study, we used an automated approach to extracting and organising all of the textual content of each firms' website. Specifically, we used the NVIVO 10 software package to identify words (or word groups) and count the frequencies with which these occurred. Note that although in most cases a variety of words that share the same stem were in practice drawn out; for simplicity, we hereafter refer to these as words, rather than words or word groups.<sup>1</sup> Generally, this involved creating an Adobe Acrobat file for the whole website and all the information disclosed on it, including annual reports, brochures, blogs records, corporate responsibility reports, and so on. For 73 websites this could not be done due to website restrictions. For 51 of these, we were able to use the NCapture function of NVIVO 10, which allows the pages to be extracted one by one as well as any uploaded PDF files. All of this information was then placed in an electronic folder. For 23 websites the extraction of information wasn't possible at all and we had to exclude these cases.

All websites were visited and 'mined' between the 15<sup>th</sup> of October and 20<sup>th</sup> of November 2014.

We assume that knowledge types are latent variables which cannot be observed directly. We also assume that the presence (and frequent use) of certain words is indicative of - or symptomatic of - the presence of a certain knowledge type (and its strength). Furthermore, we follow an a priori approach (Neuendorf, 2007), in which the researchers first identify the set of candidate words considered to be associated with each knowledge type.<sup>2</sup> In this, we followed Pina and Tether (2015) who previously identified sets of words associated with 'synthetic', 'analytical' and 'symbolic' knowledge. Note, however, that their analysis was only successful with regard to the latter two types. The three sets of words initially associated with each 'type' of knowledge are listed in Table 2.

---

<sup>1</sup> For example, the word innovation, would be drawn from innova\* and include innovate, innovation, innovations, innovator, etc.

<sup>2</sup> An alternative approach would be to allow the computer to hunt for co-occurring clusters of words that are not predefined.

Further to this, we sought to identify a fourth ‘type of knowledge’; that associated with understanding laws, regulations, and compliance with these. Hereafter, we label this ‘compliance knowledge’. To identify this ‘type of knowledge’, the two authors followed the procedure in Pina and Tether (2015), searching through the 800 most frequently occurring words that had been extracted to identify those considered to be possibly associated with this ‘type of knowledge’. The list of words initially associated with compliance knowledge is also reported in Table 2.

-----  
Insert Table 2 about here  
-----

Having identified 64 words considered to be related to the different ‘types of knowledge’ (between 15 and 17 for each ‘type’), the next step involved running a word search query in NVIVO 10 to count the occurrences of these specific words on each firm’s website; the resulting matrix had companies (as cases) in the rows and the frequencies of various words/ word groups in the columns.

We were interested in identifying ‘knowledge types’, rather than the presence of particular words. While acknowledging that any particular firm can be engaged in more than one ‘type of knowledge’, the identification of individual knowledge types will be harder when there is either too much information (i.e., all, or nearly all words appear at least once in many websites), or when too little information is available (i.e., few if any of the words appear at least once on websites). In fact, the number of words identified on each company’s website varied enormously. Furthermore, the occurrence of the words (or word groups) varied widely, with some words being found on 93% of the websites, while others were found on just 2%.

### **Factor Analysis: Identifying the words associated with each type of knowledge**

To examine the extent to which the identified words can be considered to be ‘symptoms’ of an underlying latent variable (i.e. a ‘knowledge type’), we undertook factor analysis on the

counts of the 64 words. Note that to reduce over-dispersion, we first took the square root of the observed count of each word within each case.<sup>3</sup>

We did this by first dividing the dataset into two randomly selected halves. Then, because the presence of words – in total and in the variety of words present – varied widely, we repeated the exercise, dividing the sample between cases where the total number of words was higher or lower than the median. In all cases, we selected only cases (i.e., companies) with a non-zero cell count of at least five (i.e., at least five of the words were found at least once) and a minimum median count of at least two. This provided a total dataset of 401 cases.

All of the analyses proceeded in the following way: first we included all the words in the factor analysis (FA), and examined the communalities; excluding words with low communalities after extraction (i.e., below 0.3). We specified Direct Oblimin rotation, an oblique factor rotation, because the presence of one ‘knowledge type’ need not be wholly independent of (and therefore orthogonal to) another. In other words, we should not expect or force the knowledge types to be orthogonal.

Generally these factor analyses found many (between 8 and 12) factors with eigenvalues greater than 1, however the strength of these typically declined rapidly, and an inspection of the scree slopes indicated that a three to five factor solution was the most likely to be valid. Following each estimation we inspected the pattern matrix to identify strong and weak loadings, deleting words with low loadings on any factor, or high cross loadings, before re-running the analysis. This gradually reduced the number of words in the analysis, until only those with high loadings on a single factor were retained.

After doing this exercise for all four sub-samples, we identified the words that loaded consistently and strongly on one factor. This resulted in the identification of 34 words,<sup>4</sup> which we included in a final factor analysis (with Direct Oblimin rotation). We again examined the communalities, and removed words where these were low, before examining the loadings and cross-loadings. We also experimented with the number factors extracted, before finally

---

<sup>3</sup> We also repeated the exercise using log transformed the data (i.e., the  $\ln(x + 1)$ , approach), where  $x$  is the count of words. The results were very similar.

<sup>4</sup> 'data', 'tribunal', 'interface', 'regulation', 'evaluation', 'models', 'grievance', 'compliance', 'creative', 'tools', 'idea', 'customised', 'network', 'optimisation', 'inspiration', 'computing', 'feel', 'legal', 'advice-advising', 'indemnity', 'consultation', 'art', 'insight', 'statutory', 'report', 'culture', 'love', 'simulation', 'research', 'designer', 'professional', 'certification', 'analytics', 'identity'.

settling on a three factor solution, with high loadings and low cross loadings. This is shown in Table 3.

-----  
Insert Table 3a and 3b about here  
-----

Each of these factors is strong, having an Eigenvalue greater than 2; the next largest factor had an Eigenvalue  $<1$ . Each is also comprised of at least seven items, with loadings  $>0.48$ . Subsequent estimation of the Cronbach's Alphas found these to be high, and above the commonly accepted threshold. Specifically, these were 0.870 (with 7 items), 0.896 (9 items) and 0.878 (7 items) respectively.

Table 3 also reports the initial categorisation of the words associated with each factor. All seven of the first factor's words were initially selected as "analytical", and all nine of the second factor's words were initially selected as "symbolic", showing strong alignment with our a priori conceptualisation. Five of the seven words in the third factor were initially identified as "compliance", with two being "synthetic"; this indicates the revealed 'knowledge type' differs slightly from that in our a priori conceptualisation. Note that, like Pina and Tether (2015), we were unable to successfully identify a set of words associated with 'synthetic knowledge'. We label the revealed factors as 'Analytical', 'Symbolic' and 'Compliance' respectively.

### **Identifying the presence and strength of firms' 'Knowledge Type(s)'**

The next challenge involved estimating the presence, strength or prominence of knowledge types for each firm according to the information provided on their websites. While the factor analysis had identified three sets of words associated with each of 'analytical', 'symbolic' and 'compliance' knowledge, using this information to identify the presence and strength of these 'knowledge types' in particular firms is a non-trivial task because the information on a company websites is not consistent. Firms choose both how much and what information to provide: some provide a lot of information, others very little. There are problems associated with both too much and too little information. Websites that provide lots of information increase the possibility of 'false positives'. That is the information on the website will

provide many symptoms of a ‘knowledge type’ even if this is not significant to the firm. On the other hand, websites which provide little information increase the possibility of ‘false negatives’; because the evidence is scarce or not present, we may conclude that a ‘knowledge type’ is not there, even if it is indeed important to the firm.

Our approach to identifying and estimating the prominence of ‘knowledge types’ therefore seeks to measure both the appearance of, and the relative strength of, the aggregated evidence from the different sets of words that appear on firms’ websites. We do this in two stages. First, we seek evidence for the presence of a ‘knowledge type’ regardless of its strength. This is done by first estimating, for each word associated with each ‘types of knowledge’, the probability that that word will occur on a particular website, regardless of its frequency. Specifically, we estimated binary logistic regressions to find  $p(w_{i,x})$ , the estimated probability that word  $w$  within the set  $x$  arises in the website of firm  $i$ .

For each word, we used the following predictors: (1) the count of occurrences (regardless of frequency) of 25 ‘control group’ words (in set  $m$ ) which were not closely associated with any of the three ‘types of knowledge’ being examined here. This group of words was largely those that we had initially expected to be associated with ‘synthetic knowledge’ but which were never identified as a coherent set of words. To this set we also added some words deleted in the early stages of the analysis outlined above; these words have low communalities or low loadings, and therefore little association with any of the identified factors; Further to this, we included (2) the aggregated sum of these 25 control group words; (3) a dummy variable for how the data was captured. As stated above, most of the data was captured by Adobe Acrobat 9 Standard, but for 51 cases this was not possible and data was captured by NCapture from NVIVO 10 instead. NCapture allows the extraction of textual content page by page which limits the amount of information mining. Consequently, in the majority of cases, data captured using NCapture functionally was smaller than data extracted from Adobe Acrobat 9 Standard. The inclusion of the dummy variable for data captured by NCapture was intended to adjust for any systematic bias arising due to this different method.

The basic intuition behind this approach is that the larger a website (in terms of written content) the more likely it is to contain any word,<sup>5</sup> and therefore greater significance should be attached to the presence of any word that occurs on a ‘small website’ relative to the

---

<sup>5</sup> Just as War and Peace is much more likely to contain any particular word than Animal Farm, because the former is much longer! With over ½ million words compared to c.30,000.

appearance of the same word on a large website. The set of twenty five control groups' words is used as proxy for website size, both in terms of the volume and variety of words used. This approach therefore seeks to controls for both the relative frequency of words, and for the size of the websites within which words are found.

A related challenge is whether to use simple word counts, or to transform the counts to reduce over-dispersion. We experimented with simple counts, taking the square root of counts (a common approach when dealing with over-dispersion in count data) and taking their natural logs, for both the count of different words arising, and for the aggregated count of control group words. Ultimately, we found that models that used square root transformations tended to produce the best fitting models, although the differences between the predicted values of various approaches were marginal.<sup>6</sup>

Having found the predicted probability that a word would occur on a website, we subtracted this value ( $p(w_{i,x})$ ), from unity when this word was found to occur. For example, suppose that the word 'designer' was found on a website, but was found to have an 80% chance of occurring on that firm's website. This provides a value of  $1-80\% = 0.2$ . Where a word is not found, the result was set to zero. We then sum these results for all words in the set  $x$ , where  $x$  is one of three subsets,  $j$ ,  $k$  and  $l$ , which are the 7, 9 and 7 words associated with Analytical, Symbolic and Compliance knowledge respectively.

Hence, our formulations for the presence of Analytical Knowledge, Symbolic Knowledge and Compliance knowledge are:

$$PAK_i = \sum_j [((1 - p(w_{i,j}))|N_{i,j} > 0)]_{i,j} + (0|N_{i,j} = 0)_{i,j}$$

$$PSK_i = \sum_k [((1 - p(w_{i,k}))|N_{i,k} > 0)]_{i,k} + (0|N_{i,k} = 0)_{i,k}$$

$$PCK_i = \sum_l [((1 - p(w_{i,l}))|N_{i,l} > 0)]_{i,l} + (0|N_{i,l} = 0)_{i,l}$$

The sums of the estimated mean probabilities, and therefore the average scores, are 1.04, 1.315 and 1.08 respectively. The sum of the estimated probabilities is higher for PSK<sub>i</sub> because this is based on nine words rather than seven. To make the results comparable, we 'correct' PSK<sub>i</sub> by multiplying by 7/9; i.e.:

---

<sup>6</sup> The results are not shown due to space considerations. They are available on request.

$$PSK_i^* = \frac{7}{9} \sum_k [((1 - p(w_{i,k})) | N_{i,k} > 0)]_{i,k} + (0 | N_{i,k} = 0)_{i,k}$$

This resulted in a mean value for  $PSK_i^*$  of 1.02. If  $PAK_i$ ,  $PSK_i^*$ , or  $PCK_i$  are zero, then there is no evidence for the presence of the respective ‘type of knowledge’ on the website of the firm  $i$ . This does not mean that the firm does not have or use this, just as the absence of symptoms does not mean a patient does not have a disease.

-----

Insert Figures 2, 3 and 4 about here

-----

Figures 1, 2 and 3 show the distributions of the values found for the presence of “Analytical knowledge”, “Symbolic knowledge” and “Compliance knowledge” respectively. All are highly skewed, as we would expect.

We consider that if  $PAK_i$ ,  $PSK_i^*$ , or  $PCK_i$  are less than one, then the evidence for the presence of the respective knowledge type is ‘weak’; if they are greater than one but less than two the evidence is ‘moderate’; while 2 or above is ‘strong’ (and 3 or above ‘very strong’).

To provide a sense check, we cross tabulate the presence of these ‘knowledge types’ by the sectors of activity of the firms in our dataset. Overall, and for each ‘type of knowledge’, the Chi-square test of independence between sector and knowledge types is rejected, and the results were in line with expectations. For example, Market Research firms are more likely to exhibit strong ‘analytical knowledge’ presence (and are less likely to exhibit weak ‘analytical knowledge’ presence), while Architecture, and Advertising and Design firms are both much more likely to exhibit a strong presence of ‘symbolic knowledge’, and are much less likely to exhibit weak presence of ‘symbolic knowledge’. Furthermore, engineering consultants were more likely to exhibit a strong presence of ‘compliance knowledge’. Beyond this, it is important to note that the sectoral differences are relative, not absolute. Of the 36 sector by ‘type of knowledge’ cell in the cross tabulation matrix, only one is empty (i.e., the count of cases is zero), indicating that while the extent of the presence of ‘knowledge types’ varies with sector, it is not fully explained by sector. In other words, the information on the presence of ‘types of knowledge’ complements rather than replicates information on sector of activity.

We have until now been concerned with the presence of knowledge types, and not their strength. To estimate their strength, we incorporate information about the frequency of words which appear on websites, information that we have hitherto ignored.

For each word in the sets  $x$ , we divide the square root of the observed count of this word,  $\sqrt{N_{i,x}}$ , by the median square root count of all words in the control group  $m$  that occur at least once on the firm's website (i.e.,  $(\sqrt{\overline{N_{i,m}}}|N_{i,m} > 0)$ ). Cases are excluded if fewer than three of the control group words are found on the website.<sup>7</sup> For each word with a positive count in the set  $x$ , we multiply this by its relative frequency  $(1 - p(w_{i,x}))$  and then sum this over the set. Hence, the derived measure  $SAK_i$ , estimates the “Strength of Analytical Knowledge” in firm  $i$ :

$$SAK_i = \sum_j \left[ (1 - p(w_{i,j})) \left( \frac{\sqrt{N_{i,j}}}{(\sqrt{\overline{N_{i,m}}}|N_{i,m} > 0)} \right) \right]_{i,j}$$

$SSK_i$  estimates the “Strength of Symbolic Knowledge” in firm  $i$ :

$$SSK_i = \sum_k \left[ \left( \frac{7}{9} (1 - p(w_{i,k})) \right) \left( \frac{\sqrt{N_{i,k}}}{(\sqrt{\overline{N_{i,m}}}|N_{i,m} > 0)} \right) \right]_{i,k}$$

And  $SCK_i$  estimates “Strength of Compliance Knowledge” in firm  $i$ :

$$SCK_i = \sum_l \left[ (1 - p(w_{i,l})) \left( \frac{\sqrt{N_{i,l}}}{(\sqrt{\overline{N_{i,m}}}|N_{i,m} > 0)} \right) \right]_{i,l}$$

Figures 4, 5 and 6 show results for the strength of each knowledge type.

-----  
 Insert Figures 5, 6 and 7 about here  
 -----

---

<sup>7</sup> Two alternatives were considered: first, taking the mean instead of the median. We prefer the median because it is more representative of the ‘typical’ word count and less likely to be inflated by outliers. Second, we have used the square root to reduce over-dispersion, but could instead have used a log transformation, which has a much greater dampening effect on extreme values.

To facilitate analysis, we again categorise the data, such that scores below 1 are considered “weak”, those from 1 to under 2 “moderate”, and 2 or greater “strong”. The cross-tabulation of this categorisation of strength against that for the presence, is shown for each ‘knowledge type’ in Table 4. This shows that presence and strength produce somewhat different distributions.

-----  
Insert Table 4 about here  
-----

To incorporate information on both presence and strength, we again re-categorise the data into a four group classification: Group 1: Presence and strength are both weak (<1); Group 2: Presence or strength is moderate (but neither is strong); Group 3: Presence or strength, but not both, is strong;<sup>8</sup> Group 4: Presence and strength are strong.

A simple cross tabulation of this classification by sector finds no absolute differences, but many strong relative differences. For example, Market Research firms are more likely to exhibit ‘analytical knowledge’; ‘symbolic knowledge’ is (understandably) much stronger in the creative sectors of Advertising/Design and Architecture, and weaker among the ‘non-creative’ Engineering Consulting and Market Research sectors. ‘Compliance knowledge’, on the other hand tends to be strongest among Engineering Consultants.

## 5. ANALYSIS

Having derived our categorisation of the prominence of each ‘type of knowledge’ (which combines information on presence and strength of presence), we now seek to ‘test’ the findings by regressing this categorisation against known characteristics of the firms.

Keeping the analysis simple, we utilise as independent variables three characteristics: firm size, sector of activity, and the responses to a multi-item survey question which asked about the significance of types of knowledge to the firms (although that term was not used). Specifically, the survey asked the respondents to rank the importance of items “to the

---

<sup>8</sup> Note that strength is never strong is presence is weak

services you provide to your clients”. The question and the pattern of response is shown in Table 5.

-----  
Insert Table 5 about here  
-----

The first of the items (“Analytical techniques, calculations, and/or scientific methods of evaluation”) was expected to relate to ‘Analytical Knowledge’, the third (“Inspiration, intuition and creativity”) to ‘Symbolic Knowledge’ and the second (“Knowledge of laws, statutes and/or regulations”) to ‘Compliance Knowledge’. Two further items asked about the importance of “Practical know-how, based on experience”, and the “Ability to work with other relevant experts”. The first of these was intended to identify ‘synthetic knowledge’, while the later relates to networking capabilities, and social and relational capital.

Interestingly, of all the items, firms typically rated ‘practical know-how, based on experience’ as being the most important; it was the most likely to be identified as “very important” or “crucial”. Indeed, only 96% rated this item at these levels. This may indicate that ‘synthetic knowledge’ is pervasive, and therefore not really a differentiator, being necessary rather than sufficient for competitive advantage. Perhaps because it is pervasive it is less likely to be communicated on company websites, while its strong tacit base also makes it harder to communicate.

As mentioned, the ability to work with other relevant experts relates networking capabilities and to relational and social capital, and this is also clearly important to these firms. We do not however examine it here. Also notable is that for most firms several of these types of knowledge are very important or crucial.

To analyse the data, we created two dummy variables for each of the first three items, one for when this was considered “crucial”, the other for when it was considered “very important”.

We start our modelling (using ordinal logistic regressions) – which is intended only to find associations, and should not imply causation - with only firm size (categorised) and sector of activity included. Firm size and sector have long been understood to be influential in ‘explaining’ differences in firm behaviour (e.g., Pavitt, 1984). Here, we anticipate that as

firms become larger the prominence of each knowledge type is likely to reduce. This is for two reasons: first, small firms are likely to be more specialised in terms of their substance; second, smaller firms will find it easier to develop a stronger, clearer message about their identity. In other words, they will find it easier to present a distinctive message.

With regard to sector, we anticipate significant differences between sectors due to the different activities we have included in the database. In particular, we anticipate that Market Research firms will be the most strongly oriented to ‘analytical knowledge’, followed by the Engineering Consultants. We anticipate that the Advertising/Design firms and Architecture Practices will be the most strongly oriented to ‘symbolic knowledge’; and we anticipate that Engineering (and perhaps Architecture) firms will be the most strongly oriented to ‘compliance knowledge’. Note however that we expect relative rather than absolute differences between sectors. In other words, we expect that information on sector will only partially explain the differences in the revealed prominences of the knowledge types. Not all firms in a sector share the same ‘knowledge type’ characteristics.

We undertook the modelling in three steps. In Model 1, we only include firm size and sector. In Model 2, we introduce the survey derived dummy variables for the most relevant knowledge source (i.e., for the ‘analytical knowledge’ regression, we add the dummy variables for the Survey Item 1; for the ‘symbolic knowledge’ regression, we add the dummies for Survey Item 3; and for the ‘compliance knowledge’ regression we add the dummies for Survey Item 2). In the third and final model, we add the dummy variables for all three survey items to the models for each ‘type of knowledge’. The frequency table is provided in Table 6. We will report the results for each “knowledge type” in sequence; the regression results are provided in Tables 7, 8 and 9 respectively

-----

Insert Tables 6, 7, 8 and 9 about here

-----

### **Modelling the prominence of ‘Analytical Knowledge’**

Our first model shows that, as expected, ‘analytical knowledge’ – as indicated by our website analysis – is much more prominent among Market Research firms than firms in any other sector, including Engineering Consultants and even more so Architecture and Advertising/Design firms. Secondly, the prominence of analytical knowledge tends to decline with firm size. Medium sized (50 to 249 employees) and Large firms (250+ employees) are significantly less likely to prominently exhibit ‘analytical knowledge’ on their websites than are Small firms. All of these findings are in line with expectations.

The inclusion of the dummy variables derived from the survey response does not improve the model, as neither is significant. Indeed, oddly, firms which claimed that “Analytical techniques, calculations, and/or scientific methods of evaluation” were “very important” to their services were almost significantly (at the 10% level) less likely to display ‘analytical knowledge’ on their websites. The coefficient on this being “crucial” is negative but insignificant. These findings are not in line with our expectations.

Interestingly, however, when the survey information from the other two items was included, these were found to be significant. Specifically, firms which stated that “Inspiration, intuition and creativity” were “crucial” or “very important” were significantly less likely to prominently display ‘analytical knowledge’ on their website, while those that stated “Knowledge of laws, statutes and/or regulations” was “crucial” or “very important” were significantly more likely to prominently display ‘analytical knowledge’ on their website. The aversion to ‘inspiration, etc.’ is understandable, while the affiliation to ‘knowledge of laws’ is not. This suggests that there may be a misalignment between our conceptualisation of ‘analytical knowledge’ (as displayed by company websites), and that in the minds of the survey respondents.

Overall, the models for ‘analytical knowledge’ were statistically significant, but rather weak, the best having a (McFadden) pseudo R-squared of 0.10.

### **Modelling the prominence of ‘Symbolic Knowledge’**

As with ‘analytical knowledge’, our first model for ‘symbolic knowledge’ displays expected features. ‘Symbolic knowledge’ (as shown on websites) is much more prominent among Advertising/Design firms and Architecture practices than among Market Research firms, but

also (and more surprisingly) significantly weaker among Engineering Consultancies than even Market Research firms. Secondly, the prominence of ‘symbolic knowledge’ tends to decline with firm size, but interestingly at a smaller size than for ‘analytical knowledge’; even Small firms (with 20 to 49 employees) are less likely to prominently exhibit ‘symbolic knowledge’ than Micro firms with fewer than 20 employees. These findings are in line with expectations.

In contrast to the inclusion of the dummy variables derived from the survey response into the ‘analytical knowledge’ model, the inclusion of the dummy variables indicating that “Inspiration, intuition and creativity” are “crucial” or “very important” to the firm are not only positive and highly significant, but furthermore the coefficient for “crucial” is the greater in magnitude than that for “very important”. This is wholly in line with expectations.

The further inclusion of the survey response variables shows that considering “knowledge of laws, etc.” to be “crucial” or “very important” had no significant relationship with the prominence of ‘symbolic knowledge’ (on the firm’s website), but that firms which considered the use of “Analytical techniques, calculations, and/or scientific methods of evaluation” to be “crucial” to their activities were significantly less likely to prominent signs of ‘symbolic knowledge’ on their website. This implies that ‘symbolic knowledge’ is further from ‘analytical knowledge’ (as understood by the survey respondents) than ‘knowledge of laws, etc.’

Also notable is that in contrast to the models for ‘analytical knowledge’, the models for ‘symbolic knowledge’ are strong, being highly significant, and with the best having a (McFadden) pseudo R-squared of 0.25.

### **Modelling the prominence of ‘Compliance Knowledge’**

Lastly, we report the findings of the modelling with respect to the prominence of ‘compliance knowledge’ (as estimated from the website information). Overall, these models are also statistically significant, but weaker than for ‘symbolic knowledge’, the best having a (McFadden) pseudo R-squared of 0.09.

Our first model shows that, as expected, ‘compliance knowledge’ is most prominent among Engineering Consultancies, followed by Architecture practices. Meanwhile, Advertising/Design firms are significantly less likely than Market Research firms (the

reference category) to exhibit this. Meanwhile, the prominence of ‘compliance knowledge’ declines with firm size, especially among Medium sized and Large firms. These findings are all in line with expectations.

As with ‘symbolic knowledge’, the inclusion of the dummy variables derived from the survey response significantly improves the model, with the both dummies for “Knowledge of laws, statutes and/or regulations” being “crucial” and “very important” being positive and statistically significant, and moreover the former is larger than the latter. These finding are again in line with expectations.

Finally, the further inclusion of the survey response variables shows that considering “analytical techniques, etc.” to be “crucial” or “very important” were found to have no significant relationship with the prominence of ‘compliance knowledge’ (on the firm’s website), but firms which considered that “Inspiration, intuition and creativity” were “crucial” or “very important” to their activities were significantly less likely to prominently display ‘compliance knowledge’ on their websites. This implies that ‘compliance knowledge’ is closer to “analytical knowledge” as understood by the respondents, than knowledge based on creativity, etc.

## **6. DISCUSSION AND CONCLUSIONS**

In recent years, researchers have developed conceptualisations of ‘types of knowledge’ which have extended beyond the familiar - and perhaps stale - discussion of ‘tacit’ and ‘codified’ knowledge’. Especially interesting is Asheim and colleagues (2005) conceptual identification of a third ‘type of knowledge’: ‘symbolic knowledge’.

The identification of ‘symbolic knowledge’ is particularly significant because it opens the door to conceiving further ‘types of knowledge’. For example, knowledge of laws, regulations and statutes – which combines ‘codified’ knowledge of the content of these, and ‘tacit’ knowledge as to how they are likely to be interpreted by judges - does not fit easily with any of the SAS (‘synthetic’, ‘analytical’ or ‘symbolic’) types. This suggests that there is another ‘type’ of knowledge: ‘compliance knowledge’. And if ‘compliance knowledge’ exists, what about other ‘types of knowledge’?

Asheim and colleagues' (2005) conceptualisation of multiple knowledge types also poses two other challenges. The first is how can we measure distinctive 'types of knowledge'? Perhaps surprisingly, in empirical work these concepts have been applied mostly at a high level of aggregation – such as to regions, or industries. Very limited work has sought to apply these concepts at the level of firms. Three exceptions are Martin (2012) and Tether et al., (2012) – both of which assume that different occupations relate to particular 'types of knowledge' – and Pina and Tether (2015), who attempted to identify knowledge types directly by reviewing websites.

The second challenge that Asheim and colleagues' (2005) conceptualisation creates is the question about the relationship between these 'types of knowledge'. Once we accept there are more than two types of knowledge, then these need not be equidistant from each other. For example, type A may be relatively close to B, and more distant from C. This is significant, because it has implications for diversification. Firms initially engaged in A should find it relatively easy to migrate into the B, while finding it more difficult to engage in C.

This paper has sought to pick up these challenges and we consider that the paper has made four principal contributions. First, it identifies a hitherto unmeasured 'type of knowledge': 'compliance knowledge', which is associated with understanding laws and regulations, and how to comply with these, and engaging with clients with respect to compliance. Not only have we conceptualised this 'type of knowledge', but we have also used empirical methods which have succeeded in identifying a set of words which provide a cluster of 'symptoms' indicating the presence of 'compliance knowledge'.

Second, and more generally, we have developed a methodology for identifying knowledge types from text based evidence. This methodology could in principle be applied to any text rich sources pertaining to firms, including annual reviews or case study materials. A major challenge with this is that the extent of information can vary enormously: there is a great deal of material about some firms, while there is very little about others. Like a patient visiting a doctor, the challenge is to separate the signal from the noise: the hypochondriac says so much it is difficult to know what is significant, while the stoic presents so little information that important symptoms can be overlooked. We addressed this by developing a method that estimates the probability that words would occur in a document (including on websites) controlling for the size of the document (by word count) and the variety of words used. We later built on this by adding in evidence of the relative strength of the words present.

Third, we applied this methodology to information drawn from company websites, demonstrating that it is practical.

Fourth, we tested our measures by deriving a classification based on the prominence of the three ‘types of knowledge’ identified. This testing incorporated the size and sector characteristics of the firms, and included some survey evidence. Most of the findings aligned with our prior expectations. For example, the prominence of all three ‘types of knowledge’ varies significantly by sector and in ways that could be anticipated. For example (and without recounting all of the findings), Market Research firms are the most likely to display ‘analytical knowledge’, Advertising/Design firms are the most likely to display ‘symbolic knowledge’, and ‘Engineering firms’ are the most likely to display ‘compliance knowledge’. Importantly, sector does not wholly explain the prominence of these ‘types of knowledge’, which therefore complement rather than replicate existing information.

Another interesting finding is that the prominence of each of these ‘types of knowledge’ declines significantly with firm size, implying that the smallest firms are the most specialised and have the purest expressed identify. Also interesting is that the prominence of ‘symbolic knowledge’ appears to decline earlier in the size distribution, which perhaps indicates that it is more difficult to scale this more personal ‘type of knowledge’. Furthermore, the addition of evidence from a survey concerning the importance of ‘inspiration, intuition and creativity’, and ‘knowledge of laws, statutes and/or regulations’ aligned with the website information for ‘symbolic knowledge’ and ‘compliance knowledge’ respectively. Finally, our analysis also indicates that these three ‘knowledge types’ are not equidistant: ‘analytical knowledge’ and ‘compliance knowledge’ are closer to one another, and more distant from ‘symbolic knowledge’.

As with all studies, ours has both limitations and wider lessons. In terms of limitations, we have examined only knowledge intensive service firms (or professional service firms) in some specific domains. We have not attempted to identify and measure all knowledge types in all sectors. Furthermore, in relation to the website information, we have only incorporated text based formation, excluding potentially informative but hard to code information relating to ‘look and feel’ and graphics. Thirdly, our research process depended in part on judgement – first in terms of the words chosen as possible ‘candidates’ to represent each ‘type of knowledge’, and secondly in terms of reducing the words included in the Factor Analysis. Factor Analysis is (almost) always part science, part art.

In terms of wider lessons, there are also several. The first is that websites do provide interesting and valuable information, but this information is not easy to analyse. Some websites contain an almost overwhelming amount of information, while others cannot be searched automatically or provide little information. Our approach tended to work best for the websites of smaller firms, which tend to be smaller. Arguably this is beneficial, as smaller firms are more difficult to research in a representative way by other means, such as surveys.

We also need to be mindful that websites do not provide a full and true picture of the firm: they are likely to emphasise instead what it is that the firm wants its audiences to know about. Our method is based on latent variables, but even so, we should be aware that firms – and particularly small firms – are likely to emphasise strongly information pertaining to the services they provide (or wish to provide) to clients; back room activities are likely to enjoy prominence, even if they are fundamentally important.

Finally, as Pina and Tether (2015), we failed to empirically identify ‘synthetic knowledge’. The results from the survey of firms perhaps indicates that ‘synthetic knowledge’ is so pervasive as not to be a distinguishing ‘type of knowledge’; it is also the most tacit, which makes identification difficult.

[Please note that this is a working paper that requires refinement. We trust that any reader who has got this far has found it interesting. We would of course welcome any comments as to how the paper could be improved.]

## REFERENCES

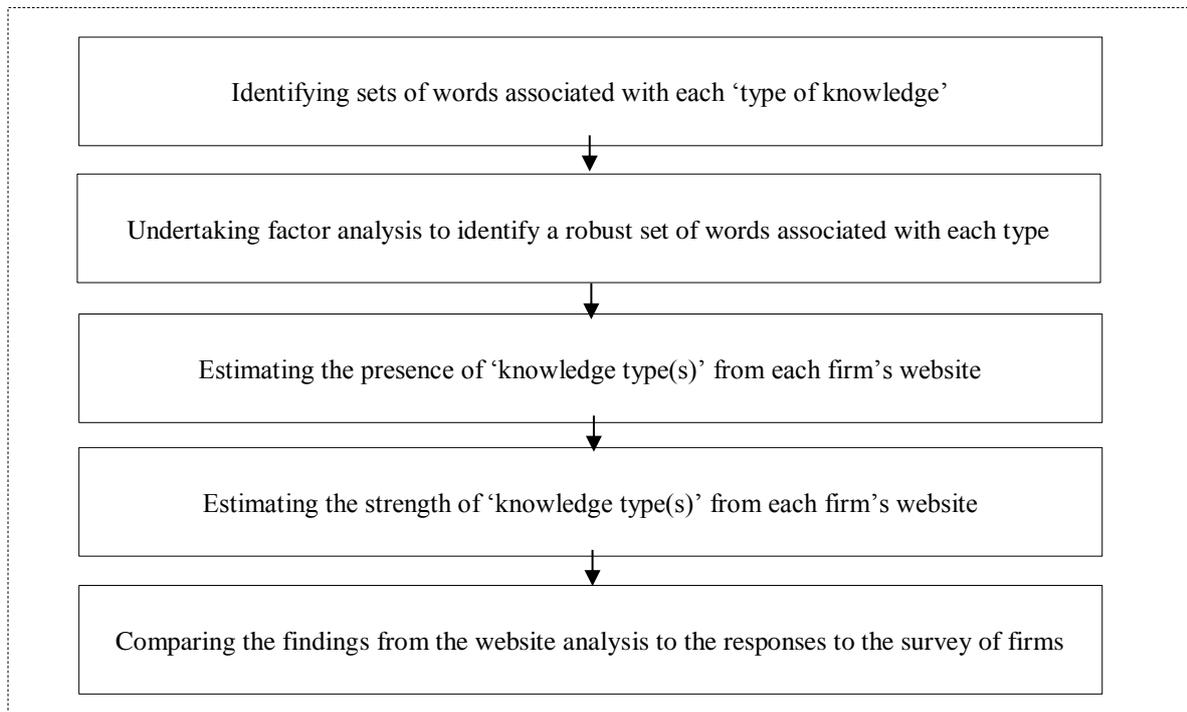
- Albert, S. and Whetten, D. (1985) Organizational identity. *Research in Organizational Behavior* 7: 263–295.
- Alvesson, M. (1990) Organization: From substance to image? *Organization Studies*, 11: 373–94.
- Asheim, B. & Coenen, L. (2005) Knowledge bases and regional innovation systems: Comparing Nordic clusters, *Research Policy*, 34:8, 1173–1190.
- Asheim, B., Coenen, L., Moodysson, J. & Vang, J. (2005) Regional Innovation System Policy: a Knowledge-based Approach. CIRCLE, Working Paper 2005/13, University of Lund, Sweden (ISSN: 1654-3149).
- Asheim, B., and Gertler, M. 2005. The geography of innovation: Regional innovation systems. In *The Oxford handbook of innovation*, ed. J. Fagerberg, D.C. Mowery, and R. R. Nelson, 291–317. Oxford, U.K.: Oxford University Press.
- Asheim, B. & Hansen, H. (2009) Knowledge bases, talents, and contexts: On the usefulness of the creative class approach in Sweden, *Economic Geography*, 85:4, 425–442.
- Baak, D. W. and Singh, N. (2007) Culture and web communications. *Journal of Business Research* 60(3): 181–188.
- Balmer, J. and Greyser, S. (2002) Managing the multiple identities of the corporation. *California Management Review*, 44(3): 72-86.
- Berelson, B. (1952) *Content Analysis in Communication Research*, Glencoe, IL: Free Press.
- Brønn, A. and Martinsen, H. (2006) A Reflective Approach to Uncovering Actual Identity”, Peggy Simic Brønn, Andreas Engell and Håvard Martinsen, publisert i *European Journal of Marketing*, 40:7/8, 886-901
- Budd, R., Thorp, R. and Donohew, L. (1967) *Content Analysis of Communications* (New York: The Macmillan Company), 2.
- Coenen, L., & Moodysson, J. (2009): Putting Constructed Regional Advantage into Swedish Practice. *European Planning Studies*, 17(4): 587-604.
- Connolly-Ahern, C. and Broadway, S. C. (2007) The importance of appearing competent: An analysis of corporate impression management strategies on the World Wide Web. *Public Relations Review* 33(3): 343–345.
- Cornelius, N., Wallace, J. and Tassabehji, R. (2007) An analysis of corporate social responsibility, corporate identity and ethics teaching in business schools. *Journal of Business Ethics* 76(1): 117–135.
- Dutton, J. E., Dukerich, J. M. and Harquail, C. V. (1994) Organizational images and member identification. *Administrative Science Quarterly* 39(2): 239–263.
- Edquist, C. (1997) *Systems of Innovation*, Pinter, London.
- Esrock, S. L. and Leichty, G. B. (1998) Social responsibility and corporate web pages: Self-presentation or agenda-setting? *Public Relations Review* 24(3): 305–319.
- Gibbons, M., Limoges, C., Nowotny, H. Schwartzan, S., Scott, P. & Trow, M. (1994) *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies*, London Sage.

- Gioia, D. A. & Thomas, J. B. (1996). Identity, Image, and Issue Interpretation: Sensemaking During Strategic Change in Academia. *Administrative Science Quarterly*, 41, 370-403.
- Ha, L. and James, E. (1998) Interactivity Reexamined: A Baseline Analysis of Early Business Web sites. *Journal of Broadcasting and Electronic Media*, 42 (2): 457-474.
- Haniffa, R. and Hudaib, M. (2007) Exploring the ethical identity of Islamic financial institutions via communication in the annual reports. *Journal of Business Ethics* 76(1): 103–122.
- He, H. W. and Balmer, J. M. T. (2007) Identity studies: Multiple perspectives and implications for corporate-level marketing. *European Journal of Marketing* 41(7/8): 22–34.
- Hwang, J., McMillan, S. and Lee, G. (2003) Corporate Web Sites as Advertising. *Journal of Interactive Advertising*, 3(2): 10-23
- Kakabadse, N., Kakabadse, A., and Kouzmin, A. (2003) Reviewing the knowledge management literature: towards a taxonomy. *Journal of Knowledge Management*, 7: 75-91.
- King, N. (2004) Using templates in the thematic analysis of text, in C.Cassell and G.Symon (Eds.) *Essential Guide to Qualitative Methods in Organizational Research*. London: Sage.
- Knox, S. and Bickerton, D. (2003) The six conventions of corporate branding. *European Journal of Marketing* 37(7/8): 998–1016.
- Kogut, B. and Zander, U. (1996) What firms do? Coordination, Identity and Learning. *Organization Science*, 7.5, pp. 502-5183: 383-397.
- Lundvall, B-A. (1993) *National Systems of Innovation*, Pinter, London
- Li, QC, Tether, B. and Mina, A. (2012). "Knowledge-bases, places, spatial configurations and the performance of knowledge-intensive professional service firms," Working Papers 9793, Imperial College, London, Imperial College Business School.
- Malerba F (2002) Sectoral systems of innovation and production, *Research Policy* 31:2, 247-264.
- Malerba F (2005) Sectoral systems: How and why innovation differs across sectors. In: Fagerberg J, Mowery DC, Nelson RR (eds) *The Oxford Handbook of Innovation*. Oxford University Press, Oxford
- Martin, R. (2012) Measuring Knowledge Bases in Swedish Regions. *European Planning Studies* 20 (9):1569-1582.
- Martin, R. (2013) Differentiated Knowledge Bases and the Nature of Innovation Networks. *European Planning Studies* 21 (9):1418-1436.
- Marston, C. and Polei, A. (2004) Corporate reporting on the Internet by German companies. *International Journal of Accounting Information Systems* 5: 285-311.
- McIver, D., Lengnick-Hall, C., Lengnick-Hall, L. and Ramachandran, I. (2012) Integrating knowledge and knowing: A framework for understanding knowledge-in-practice. *Human Resource Management Review*, 22: 86-99.
- McIver, D., Lengnick-Hall, C., Lengnick-Hall, L. and Ramachandran, I. (2013) Understanding work and knowledge management from a knowledge-in-practice perspective. *Academy of Management Review*, 38(4): 597-620.

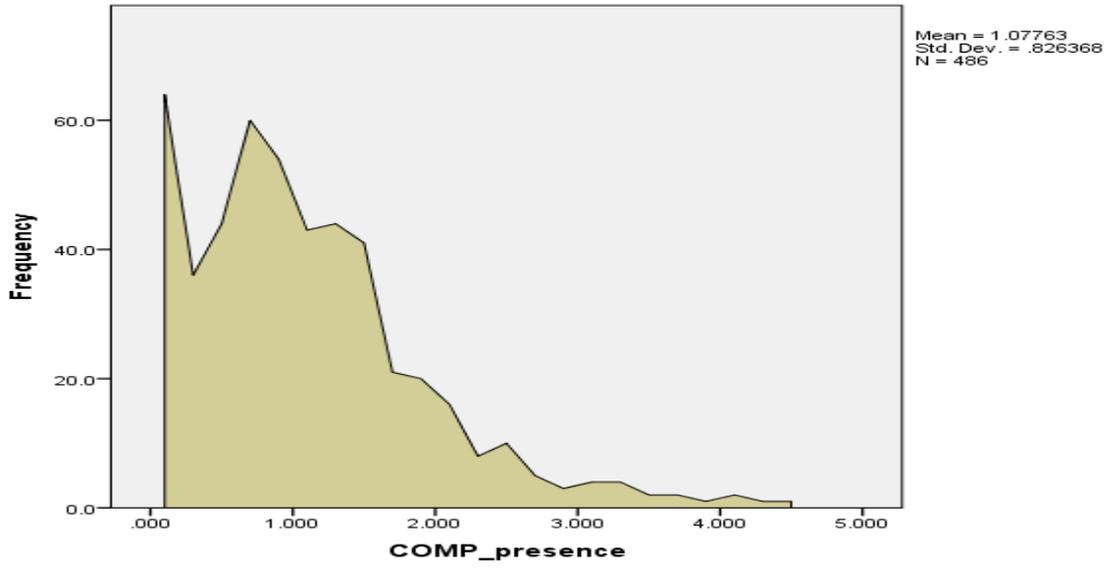
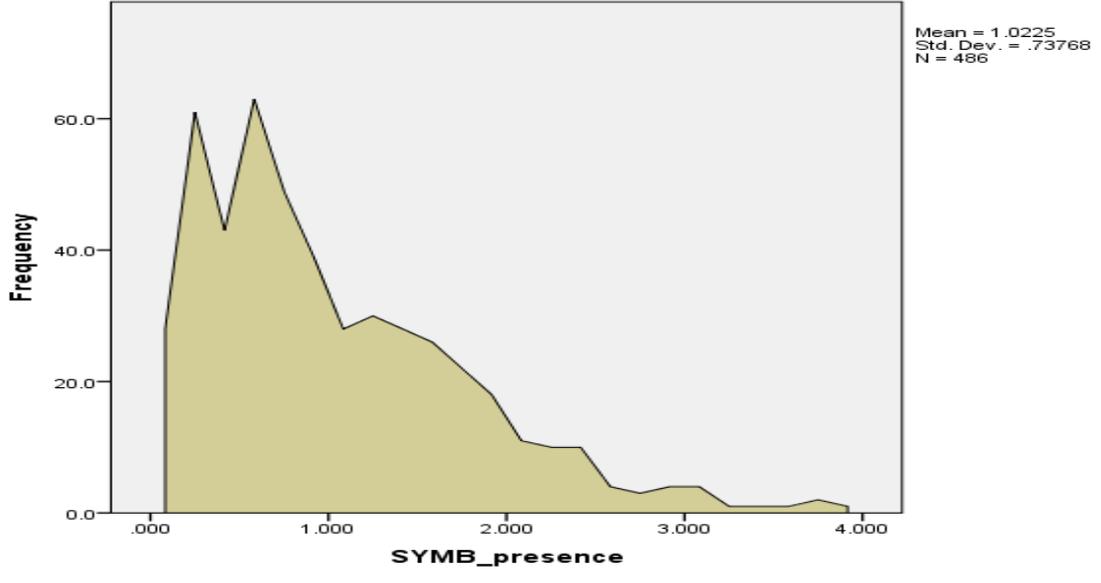
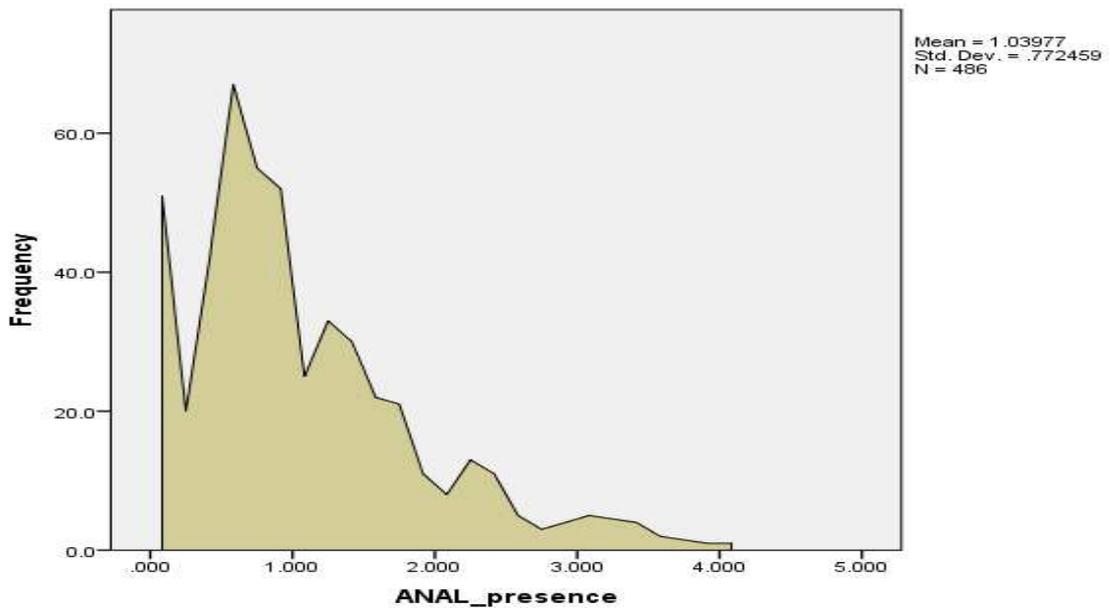
- McMillan, S. (2000) The Microscope and the Moving Target: The Challenge of Applying Content Analysis to the World Wide Web. *Journalism and Mass Communication Quarterly*, 77 (1): 80-98.
- Moodysson, J., Coenen, L. & Asheim, B. (2008) Explaining spatial patterns of innovation: Analytical and synthetic modes of knowledge creation in the Medicon Valley life science cluster, *Environment and Planning A*, 40(5), pp. 1040–1056.
- Lave, J., & Wenger, E. (1991) *Situated learning: Legitimate peripheral participation*. New York: Cambridge University Press.
- Neuendorf, K. (2002) *The content analysis guidebook*, Thousand Oaks, CA, Sage.
- Pavitt K (1984) Sectoral patterns of technical change: Towards a taxonomy and a theory. *Research Policy* 13: 343–373
- Pina, K. and Tether, B.S. (2015). Distinguishing knowledge types and the differential drivers of innovation among KIBS, Working paper, Manchester Business School.
- Polanyi, M. (1967) *The tacit dimension*, London,, Routledge & K. Paul.
- Rodrigues, S. and Child, J. (2008) The development of corporate identity: A political perspective. *Journal of Management Studies* 45(5): 886–911.
- Scott, S. & Lane, V. (2000) A Stakeholder Approach to Organizational Identity. *The Academy of Management Review*, 25, 43-62.
- STEMLER, S. 2001. An overview of content analysis. *Practical assessment, research & evaluation*, 7, 137-146.
- Stern, Barbara B. (1989), "Literary Criticism and Consumer Research," *Journal of Consumer Research*, 16 (December), 322-334.
- Tether, B.S., Li Q.C. and Mina, A. (2012) "Knowledge-bases, places, spatial configurations and the performance of knowledge-intensive professional service firms." *Journal of Economic Geography* 12.5: 969-1001
- Uzunoglu, E. and Kip, S. M. (2014) Building relationships through websites: A content analysis of Turkish environmental non-profit organizations' (NPO) websites. *Public Relations Review*, 40: 113-115

## **Figures and Tables**

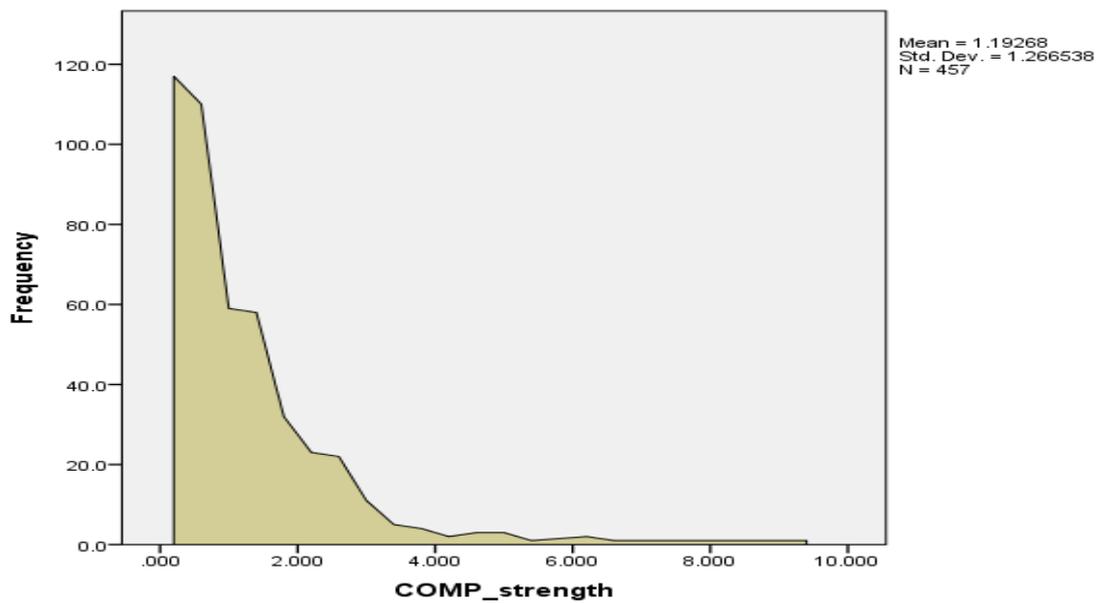
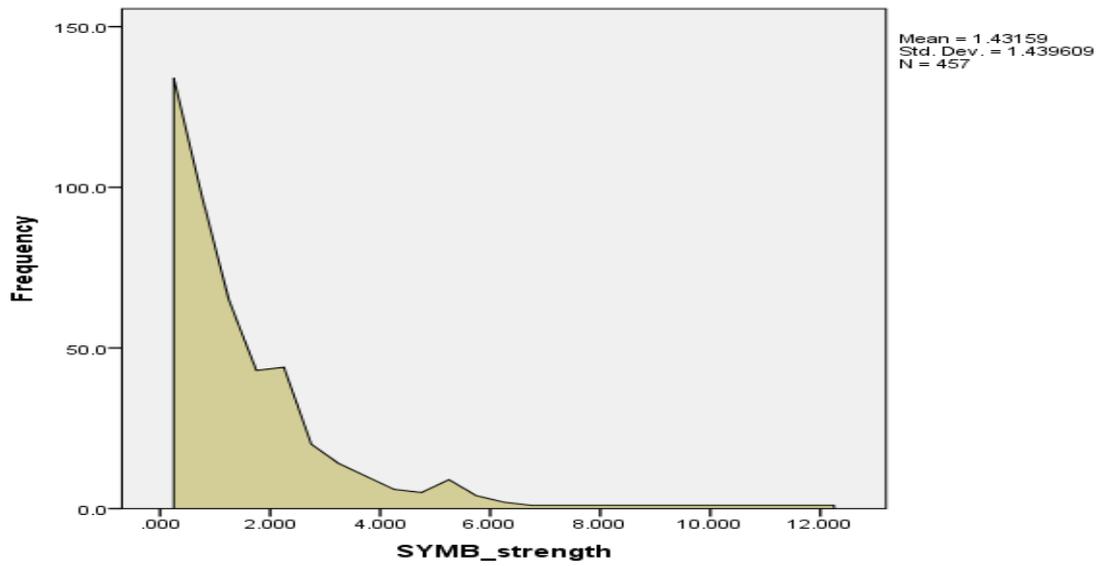
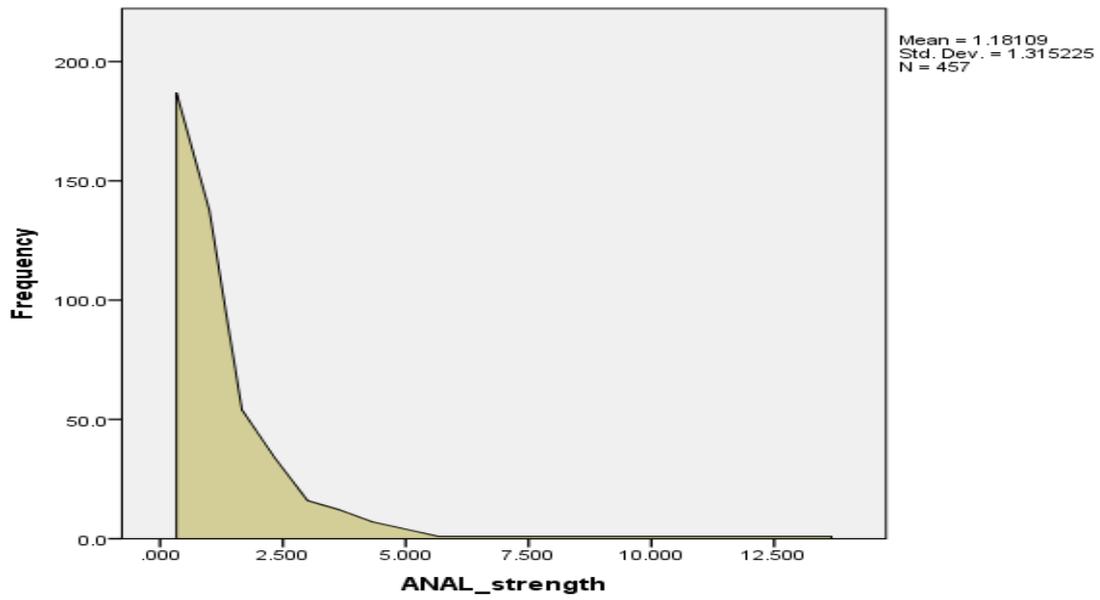
**Figure 1: Overview of the process of classifying firms by ‘knowledge type(s)’**



**Figures 2, 3 and 4 – Distributions of the Presence of the three Knowledge Types**



**Figures 5, 5 and 7 – Distributions of the Strength of the three Knowledge Types**



**Table 1: Summary Table of the SAS Types of Knowledge and their Characteristics**

	Analytical	Synthetic	Symbolic
Innovations and Solutions	Fundamental innovation by the creation of new knowledge. Solutions found by applying scientific models or equations	'Local' solutions developed by applying or combining existing knowledge. Occasionally these become general purpose 'killer applications'	Solutions based on hard to explain tacit insights. Major innovations often recognised ex post (as value is socially constructed)
Codified or Tacit?	Predominantly codified and "Scientific", based on deductive processes and formal models	Predominantly tacit and "applied, problem related". Largely practical, and often developed through inductive processes	Predominantly tacit and "Artistic". Importance of building and challenging conventions: the 'power of persuasion' matters.
Locus of new knowledge production	R&D departments and collaborations, including with the 'science base'	Interactive learning, especially with clients, but also in the community of practice	'Studio' projects, and learning through interaction with the professional/artistic community, and wider cultural interactions.
Exemplar industry	Biotechnology and other 'science based' industries (Pavitt, 1984)	'Low-tech.' engineering based industries and other 'specialist suppliers' (Pavitt, 1984)	Film directors and other 'cultural industries' (Scott, 1999)
Means of Sharing and Diffusing Knowledge	Patents, publications and the Internet, but also scientific conferences	Attending to 'field problems' (von Hippel, 1988), mainly through face-to-face interactions	Hard to share or diffuse. Developed in practice over time and 'possessed' by key individuals.
Applied to architecture & engineering consulting?	Highly analytical engineering services are here: e.g., fire and earthquake engineering. Others less so, but all have an analytical base.	Needed to get buildings built, & for different professions to coordinate. 'Low-order' architects and engineers are mainly here.	'Starchitects' are the 'high priests' of this. 'Strong idea' architectural practices are here. Strong delivery and service less so.

Adapted from Asheim et al. (2007: Table 1) and Faulconbridge et al. (2011: Table 2.5)

**Table 2: Initial set of selected words by four knowledge types**

Type	Initial set of selected words
'analytical'	<u>16 words</u> : data, tools, optimisation, models, analytics, computing, analysis, measurement, simulation, laboratory, evaluation, research, science, accurate, report and tests
'synthetic'	<u>16 words</u> : advice-advising, bespoke, collaboration, consultation, customised, discussing, engagement, implementation, interaction, interface, network, participation, partnership, procedures, solution, unique
'symbolic'	<u>15 words</u> : insight, idea, inspiration, art, studio, emotion, cultural, illustrator, feel, music, brand, identity, love, designer and creativity
'compliance'	<u>17 words</u> : assuring, certification, chartered, compliance, ethic, grievance, guarantee, indemnity, integrity, legal, liability, litigious, professional, register, regulation, statutory, tribunal

**Table 3a: Factor Analysis**

F.	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sum of Squared Loadings
	Total	% of Variance	Cum.%	Total	% of Variance	Cum.%	Total
1	9.456	41.112	41.112	8.728	37.946	37.946	6.305
2	3.203	13.927	55.039	2.986	12.984	50.930	6.401
3	2.125	9.240	64.279	1.851	8.047	58.977	6.243

**Table 3b: Distribution of words identified with each factor and their initial groupings**

Word / Word Group	Initial Grouping	Factor 1	Factor 2	Factor 3
Data	Analytical	0.978	-0.034	-0.007
Evaluate / Evaluation	Analytical	0.951	-0.100	-0.121
Analytic / Analytics	Analytical	0.623	0.020	0.008
Report	Analytical	0.605	0.146	0.073
Research	Analytical	0.578	0.157	0.039
Optimize / Optimise etc.	Analytical	0.496	0.225	0.217
Computing	Analytical	0.483	0.149	0.256
Inspiration	Symbolic	-0.088	0.867	-0.090
Creative	Symbolic	0.127	0.769	0.112
Love	Symbolic	-0.073	0.743	-0.153
Idea	Symbolic	0.170	0.735	0.084
Feel	Symbolic	0.086	0.732	0.026
Culture / Cultural	Symbolic	0.023	0.672	0.145
Art	Symbolic	0.100	0.616	0.215
Identity	Symbolic	0.062	0.601	-0.103
Designer / Designers	Symbolic	-0.023	0.575	0.239
Regulation / Regulations	Compliance	0.232	-0.108	0.841
Statutory	Compliance	-0.084	-0.002	0.787
Consultation	Synthetic	-0.029	0.146	0.735
Compliance	Compliance	0.315	-0.124	0.731
Legal	Compliance	0.188	-0.036	0.715
Advice / Advising	Synthetic	0.030	0.108	0.684
Indemnity	Compliance	-0.095	0.007	0.512

**Table 4: Cross tabulating the Presence of Knowledge Types by their Strength**

		Strength - Analytical			Strength - Symbolic			Strength - Compliance		
		Weak	Mod	Strong	Weak	Mod	Strong	Weak	Mod	Strong
Presence	Weak	230	32	6	216	32	6	215	21	3
	Moderate	40	66	26	16	71	65	49	79	31
	Strong	0	11	46	0	5	46	0	12	47
	All	270	109	78	232	108	117	264	112	81

Prominence Categorisation	Analytical	Symbolic	Compliance
1. Presence <u>and</u> Strength both Weak	230	216	215
2. Presence <u>or</u> Strength is Moderate*	138	119	149
3. Presence <u>or</u> Strength is Strong <sup>#</sup>	43	76	46
4. Presence <u>and</u> Strength both Strong	46	46	47
Total	457	457	457

\* but not strong; <sup>#</sup> Note, strength is never strong if presence is weak.

**Table 5: Survey Question and Pattern of Responses on Knowledge Used**

**Q 3 How important are the following to the services you provide to your clients?**

Please circle one number in each row.

	Of no Importance	Of minor importance	Quite Important	Very Important	Crucial
Analytical techniques, calculations and/or scientific methods of evaluation .....	5%	15%	26%	26%	29%
Knowledge of laws, statutes and/or regulations .....	3%	12%	20%	36%	29%
Inspiration, intuition and creativity .....	1%	3%	14%	32%	50%
Practical know-how, based on experience .....	0%	0%	3%	46%	50%
Ability to work with other relevant experts .....	1%	3%	16%	45%	35%

Note, number of valid responses varies from 448 to 451 among the sample examined here.

**Table 6: Frequency Table**

Variable	N.	Variable	N.
Sector: Advertising / Design	66	Analytical etc. = Crucial	128
Sector: Architecture	184	Analytical etc. = V. Impt	115
Sector: Engineering Cons.	163	Analytical etc. < V. Impt	206
Sector: Market Research	42	Inspiration etc. = Crucial	225
Size: 250+ employees	38	Inspiration etc. = V. Impt	144
Size: 50 to 249 employees	74	Inspiration etc. < V. Impt	82
Size: 20 to 49 employees	120	Laws etc. = Crucial	132
Size: 10 to 19 employees	110	Laws etc. = V. Impt	160
Size: 1 to 9 employees	113	Laws etc. < V. Impt	156

**Table 7: Ordinal Logistic Regression – “Analytical Knowledge”**

Thresholds and Variables	Model 1		Model 2		Model 3	
	B	Sig.	B	Sig.	B.	Sig.
Threshold 1 (1 to 2)	-2.551	0.000	-2.760	0.000	-3.101	0.000
Threshold 2 (2 to 3)	-0.910	0.008	-1.101	0.004	-1.409	0.001
Threshold 3 (3 to 4)	-0.026	0.941	-.196	0.606	-0.469	0.258
Sector: Advertising / Design	-2.421	0.000	-2.579	0.000	-2.310	0.000
Sector: Architecture	-2.628	0.000	-2.705	0.000	-3.026	0.000
Sector: Engineering Cons.	-1.925	0.000	-1.905	0.000	-2.319	0.000
Size: 250+ employees	-2.091	0.000	-2.172	0.000	-2.238	0.000
Size: 50 to 249 employees	-1.000	0.001	-0.998	0.001	-0.897	0.004
Size: 20 to 49 employees	-0.220	0.386	-0.221	0.390	-0.197	0.454
Size: 10 to 19 employees	-0.177	0.487	-0.238	0.356	-0.216	0.407
Analytical etc. = Crucial	Not included		-0.108	0.673	-0.156	0.577
Analytical etc. = V. Impt	Not included		-0.398	0.103	-0.377	0.134
Inspiration etc. = Crucial	Not included		Not included		-0.637	0.036
Inspiration etc. = V. Impt	Not included		Not included		-0.633	0.029
Laws etc. = Crucial	Not included		Not included		0.675	0.029
Laws etc. = V. Impt	Not included		Not included		0.613	0.029
N. Observations	455		447		441	
Model Chi-squared	87.9 (7 df, Sig. 1%)		91.3 (9 df, Sig. 1%)		102.5 (13 df, Sig. 1%)	
McFadden pseudo R-squared	0.083		0.088		0.100	

Reference sector is market research. Reference size band is under 10 employees. Reference importance of analytical techniques, etc. is “quite important” or less.

**Table 8: Ordinal Logistic Regression – “Symbolic Knowledge”**

Thresholds and Variables	Model 1		Model 2		Model 3	
	B	Sig.	B	B	Sig.	B
Threshold 1 (1 to 2)	0.183	0.618	0.984	0.040	0.829	0.099
Threshold 2 (2 to 3)	1.954	0.000	2.756	0.000	2.646	0.000
Threshold 3 (3 to 4)	3.387	0.000	4.197	0.000	4.103	0.000
Sector: Advertising / Design	2.437	0.000	1.971	0.000	1.761	0.000
Sector: Architecture	1.922	0.000	1.584	0.000	1.445	0.001
Sector: Engineering Cons.	-1.140	0.004	-1.166	0.005	-1.041	0.017
Size: 250+ employees	-2.863	0.000	-2.872	0.000	-2.882	0.000
Size: 50 to 249 employees	-0.606	0.064	-0.716	0.033	-0.746	0.030
Size: 20 to 49 employees	-0.450	0.092	-0.389	0.156	-0.267	0.337
Size: 10 to 19 employees	-0.150	0.563	-0.128	0.627	-0.041	0.877
Analytical etc. = Crucial	Not included		Not included		-0.649	0.048
Analytical etc. = V. Impt	Not included		Not included		0.026	0.919
Inspiration etc. = Crucial	Not included		1.296	0.002	1.445	0.001
Inspiration etc. = V. Impt	Not included		0.930	0.026	0.955	0.023
Laws etc. = Crucial	Not included		Not included		-0.068	0.841
Laws etc. = V. Impt	Not included		Not included		-0.100	0.746
N. Observations	455		449		441	
Model Chi-squared	251.5 (7 d.f., Sig. 1%)		261.2 (9 d.f., Sig. 1%)		266.9 (9 d.f., Sig. 1%)	
McFadden pseudo R-squared	0.224		0.236		0.246	

Reference sector and size is as before. Reference importance of inspiration etc. is “quite important” or less.

**Table 9: Ordinal Logistic Regression – “Compliance Knowledge”**

Thresholds and Variables	Model 1		Model 2		Model 3	
	B	Sig.	B	Sig.	B	Sig.
Threshold 1 (1 to 2)	0.301	0.410	0.481	0.199	-0.071	0.867
Threshold 2 (2 to 3)	1.939	0.000	2.145	0.000	1.612	0.000
Threshold 3 (3 to 4)	2.783	0.000	3.009	0.000	2.493	0.000
Sector: Advertising / Design	-0.920	0.039	-0.851	0.063	-0.757	0.127
Sector: Architecture	0.718	0.040	0.139	0.722	0.243	0.568
Sector: Engineering Cons.	1.487	0.000	1.045	0.006	1.022	0.008
Size: 250+ employees	-1.229	0.002	-1.423	0.000	-1.355	0.001
Size: 50 to 249 employees	-0.590	0.050	-0.596	0.052	-0.502	0.107
Size: 20 to 49 employees	-0.153	0.551	-0.233	0.378	-0.200	0.456
Size: 10 to 19 employees	-0.177	0.494	-0.189	0.472	-0.163	0.539
Analytical etc. = Crucial	Not included		Not included		-0.266	0.334
Analytical etc. = V. Impt	Not included		Not included		-0.266	0.282
Inspiration etc. = Crucial	Not included		Not included		-0.635	0.034
Inspiration etc. = V. Impt	Not included		Not included		-0.664	0.019
Laws etc. = Crucial	Not included		1.185	0.000	1.335	0.000
Laws etc. = V. Impt	Not included		0.706	0.008	0.724	0.008
N. Observations	455		446		441	
Model Chi-squared	66.7 (7 d.f., Sig. 1%)		84.1 (9 d.f., Sig. 1%)		92.0 (9 d.f., Sig. 1%)	
McFadden pseudo R-squared	0.062		0.079		0.087	

Reference sector and size as before. Reference importance of knowledge of laws is “quite important” or less.