Nesta...

creative skillset
Developing world class talent

MODEL WORKERS

How leading companies are recruiting and managing their data talent

Hasan Bakhshi, Juan Mateos-Garcia and Andrew Whitby. July 2014

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Co-Investment from
UKCES
UK COMMISSION FOR EMPLOYMENT AND SKILLS
Employer Investment Fund

Acknowledgements

We would like to thank the members of the advisory board for this project: **Kion Ahadi, Matthew Bayfield, Francine Bennett, Naomi Gummer, Laura Hart, Andrew Harvey-Price, Peter Hounsome, Tom Khabaza, Hamish Nisbet, Christian Reimsbach-Kounatze, Ammon Salter, Philippe Schneider, Ralph Schroeder, Rachel Schutt, Hetan Shah, Jeremy Silver, Ed Steinmueller, Greg Taylor, Richard Vidgen** and **Chris Yiu.**

We would also like to thank **Liz Murphy, Jolyon Fairburn-Beech, Jan Shury** and **David Vivian** at IFF for their research support.

Nesta...

Nesta is the UK's innovation foundation.

An independent charity, we help people and organisations bring great ideas to life. We do this by providing investments and grants and mobilising research, networks and skills.

Nesta is a registered charity in England and Wales with company number 7706036 and charity number 1144091. Registered as a charity in Scotland number SCO42833. Registered office: 1 Plough Place, London, EC4A 1DE.

www.nesta.org.uk

MODEL WORKERS

How leading companies are recruiting and managing their data talent

Contents

Executive Summary

This report draws on 45 interviews with industry experts working at the coalface of the UK's data economy to explore the following questions:

- What are the data analysis skills needs of leading companies?
- How do these companies find, manage and retain their data talent?
- How do policy and education need to change to improve the supply of data talent in the UK?

I. Context: UK companies across all sectors are experiencing a quantitative turn

- There isn't a 'one size fits all' for data analysis. The companies we have talked to can be classified into four broad 'data modes' (Business Intelligence, Data–Intensive Science, Web Analytics and Big Data), depending on the data sources they use and how they apply their data.
- Big Data companies – the newest and arguably most innovative data mode – look for valuable patterns in large and messy data sets, use those patterns to automate and predict, and design systems (machines) able to learn over time.
- Data scientists have the hybrid skill set to do Big Data work. Although statistics is at the heart of what they do, they also need programming skills to access and manage data in the first place, and to create data products and services.
- Big Data companies are in the minority in our sample – they are a quarter of the companies we interviewed – but this is changing, as more companies in other data modes start experimenting with new and bigger data. For example, Business Intelligence companies are moving from description to prediction, Data–Intensive Science Companies are exploring new web data sources, and Web Analytics companies are developing automated products and services.
- This experimentation is driven by supply–side factors (more data and better technologies to deal with it) and demand–side factors (competitors are doing it, and customers are demanding faster, more personalised services). As a result there is strong demand for new types of data talent.

II. Our interviewees report a severe shortage of UK data talent with the right skills

- Almost all the companies we interviewed, regardless of their data mode, are looking for analysts with a data scientist profile, including a mix of analytical and coding skills, and creativity and business know–how.
- This convergence of data skills needs is driven by the increasing availability of bigger and messier data in all sectors of the economy.
- Four in five of the companies we interviewed are, however, struggling to find the talent they need. The skills shortages appear to be even more severe outside of London.
- Companies give several reasons for these skills shortages:
  - Data talent is in short supply: there are too few candidates with the skills demanded.
  - Data talent in the market lacks the right skills and experience: seasoned data analysts are very expensive, and junior people require extensive training.
  - Data talent lacks the right mix of skills: good analysts often cannot code, and good coders often cannot analyse. Data analysts with the commercial nous to create business impacts are very rare.
  - Many companies lack the capacity to recruit data talent effectively. There are hurdles in communicating data candidate specifications to HR managers and recruitment agencies, assessing the skills of candidates, and understanding the business value they can create.

III. The long-term outlook for the market for data talent is uncertain

- Some companies think that changes in education and new technologies (such as better tools for data analysis) will alleviate the pressures on the market for data talent, while others think that competition for data talent between sectors will heat up the market further.
- Some companies are considering – or have already started – offshoring their data analysis capabilities outside of the UK.

IV. Policymakers and educators need to take urgent action to address the near-term problems

In the near term, action is needed to:

1. Develop the skills of the existing UK data workforce through targeted training and Continuous Professional Development (CPD).
2. Build up the ‘data analyst’ profession to develop training standards, liaise with education and raise awareness about the value of data analysis in business.

3. Ensure UK employers can recruit overseas data talent to remedy current skills gaps.
4. Create better linkages between employers and universities.
5. Improve the supply of data talent with the right mix of skills from education.

In the longer term, the UK needs to:

6. Strengthen its education system to improve numeracy and data handling across the board.
7. Change perceptions of data analysis as boring and uncreative.

V. Managers can learn from the leaders in adopting good practices

- Data analysis work is often innovative: the rewards can be great, but so are the management challenges in achieving them (and the risks of failure). The companies we interviewed are developing good practices and strategies to overcome these challenges.

Action	Practice
1. Finding data talent	• **Building a reputation** for doing interesting data work by publicising data analyses, participating in conferences and meet-ups, and contributing open data and source code back to the community. • **Going where the talent is,** including online communities and competitions. • **Working with universities** to identify promising new talent. • Always **testing candidates for their technical skills.**
2. Building data teams	• **Creating balanced teams;** does data analysis inside the company require specialists or generalists? • **Creating diversity by design,** because teams with varied skills and mindsets are more flexible, innovative and able to learn. • **Developing a shared language** to make communication more efficient.
3. Placing talent in the right part of the organisation	• Setting up central data teams helps build a critical mass of data skills, but there is a risk these teams can become ivory towers. This can be avoided by **creating strong communication channels with the rest of the organisation.** • Embedding data talent in other business functions improves communications, but can result in silos. This can be avoided by **creating data talent communities spanning the organisation.**
4. Data playing, and doing	• **Empowering data talent:** Creative data analysts are most productive when working in interesting projects over which they have a feeling of ownership. • At the same time, it is important to **channel talent into activities that create business value,** preventing data analysts from going down analytical rabbit holes where there is little business value. • Failure is a reality of all innovative projects – including data analysis work. This **risk of failure has to be accepted and managed.**

1. Introduction

The human face of the data revolution

Every day brings a new story about the explosive pace at which data is being created, and about the business opportunities and threats that this represents.[1]

Many of these stories focus on the raw material – the high volume, variety and velocity data sets that businesses can now access – and on outputs – the services and platforms built with this data. The critical stage in the middle – the actions that transform data into value for users and shareholders – receives much less attention.

Yet, we know that amassing data on its own does little for the bottom line – data needs to be analysed, and the insights applied, and this involves skilled workers.[2] Algorithms might well be the commercial workhorses of the data economy, but they are created by human talent.

What are the skills and competences that leading companies look for in this data talent? Where do they find it? How do they manage it? Will the education system be able to supply it?

These are the questions that we address in 'Skills of the Datavores', Nesta's programme of research into the workforce skills implications of the data economy that we are running in partnership with Creative Skillset and the Royal Statistical Society.[3] This report is the first output from that research programme.

The context for our research

Analysts have always flocked to data–rich industries, and new professions and academic disciplines such as actuarial science, operations research, financial mathematics and epidemiology have emerged to deal with new sources of industry data.[4]

This is happening again. Innovations in software, hardware and networks have given rise to new and big data and also, in parallel, a new profession – 'the data scientist' – to analyse it.[5] In most definitions, data scientists have programming and database skills to access and 'wrangle' unruly data, statistical skills to extract insights from it, and the business knowledge to transform those insights into impacts.[6] A growing library of consultancy reports, executive surveys and business anecdotes suggests that people with these skills are in short supply.[7]

However, not everyone is sold on the idea of data scientists. Some say that looking for them is like 'chasing unicorns', because people able to master all those skills are so rare.[8] Others complain that the data scientist label has become too vague a description of an occupation to be useful in practice.[9] There is also scepticism about whether a data scientist is a novel occupation, or rather a rebadging of existing ones like 'statistician' or 'data miner'.[10] Others still question the extent to which data science is relevant to industry in general as compared to a small elite of the most data–intensive companies.[11]

There is a risk that confusion around skills might prevent businesses from exploiting new data opportunities. It makes it difficult for employers to recruit the right talent, and for talent to find what jobs to apply for. It makes it challenging to communicate the value of data talent inside businesses. And if the relationships between new occupations like 'data scientist' and existing disciplines are unclear, it makes it harder for educators to design programmes that address the needs of industry, and for businesses to know who to speak to in universities.

About this report

We have interviewed 45 experts involved in recruiting and managing data talent with the aim of clarifying these issues. We want to understand what specific skills businesses want from their data talent, identify good practices when it comes to recruiting and managing them, and determine how policy and the education system must adapt to improve the supply of data talent in the UK (see Box 1 for an overview of our research methodology).

In a companion report, to be published later this year, we continue studying these issues through the lens of a sample survey of UK firms. There, we will also look at the link between data talent recruitment and management practices, and financial measures of business performance.

Structure

Section 2.
Looks at how the companies we interviewed use data, and what this means for their data analysis skills needs.

Section 3.
Considers the UK labour market for data talent: the profile of the 'perfect data analyst' that businesses are seeking, and the talent that is available in the market. It also considers the future evolution of the labour market for data talent.

Section 4.
Presents the implications for policy and education.

Section 5.
Presents the implications for business managers.

Box 1: An overview of the methodology

This report is based on 45 interviews with industry experts including managers of data teams, data scientists and HR managers. The sample was designed to include companies that use data intensively and could therefore be expected to employ data talent. We considered seven sectors (Creative Media, ICT, Financial Services, Manufacturing, Pharmaceuticals, Retail and ICT). Most of our interviewees worked in large companies, although we also had a small proportion of SMEs and three individuals (two independent consultants and one investor).

The interviews were conducted by two teams (IFF Research and Nesta) working in coordination. IFF undertook 30 interviews, and Nesta did 15. The interviews were semi-structured, and covered topics including the analysis of data inside business, the skills needed to do so, how data talent is recruited and managed, the existence (or not) of skills shortages and potential policy actions. Two-thirds of the interviews were conducted over the telephone, and one-third face-to-face. On average, the interviews took one hour to complete.

SECTOR	
Creative media	**9**
ICT	**8**
Pharmaceuticals	**7**
Financial services	**6**
Manufacturing	**6**
Retail	**6**
Other	**3**

SIZE	
Large (>250)	**25**
Medium (50–249)	**8**
Small (<50)	**9**
Individual	**3**

Scope and terminology

Our main focus in this report is not on engineers who develop and maintain a company's data infrastructure (and who often work in IT departments), as important as this function is for collecting and making data available to others in the business.[12] Rather, we focus on those analysts whose job is to extract insights from that data – when talking about them in general, we use the terms 'data talent' or 'data analyst' interchangeably. When referring to what they do, we use the term 'data analysis'.

2. Data frame: how are UK businesses analysing their data, and what does this mean for their skills needs?

I. Data modes

There is no data 'one size fits all'

We have found great variation in how data is used and, correspondingly, in the data skills make-up of the existing workforce across the companies that we interviewed.

This is not simply caused by differences in their analytical sophistication, but also in the types of data that matter to them, and the business models they use to create value from data. So, for example, the methods to extract insights from social media differ from those needed to optimise a logistics operation. Designing clinical trials to test the effect of drugs over months or years requires different skill sets from designing algorithms to assess a lending prospect in the blink of an eye. It is important to avoid thinking of data analysis as a 'monolith' that looks the same in every organisation and sector.

We have classified the companies that we interviewed into four data modes depending on what are now three standard features of the data they use[13] (Table 1):

- **Volume: How big is the data?** This refers to the volume of the data sets that are analysed. Smaller data sets can be analysed in a single conventional computer, while larger data sets might require dedicated clusters of such computers, or high-end parallel processing.
- **Variety: How varied is the data?** Variety refers to the number of data sources that are combined for analysis, and how structured and clean the data is. Structured data is often available in 'flat' tables, and ready for analysis. Unstructured data like text, video or audio needs to be processed to extract information from it before the analysis even begins.
- **Velocity: How quickly is the data accessed and analysed?** Some companies need to analyse the data they capture instantly ('online') to deliver services like online banking or targeted advertising, while others do this periodically (in 'batches').

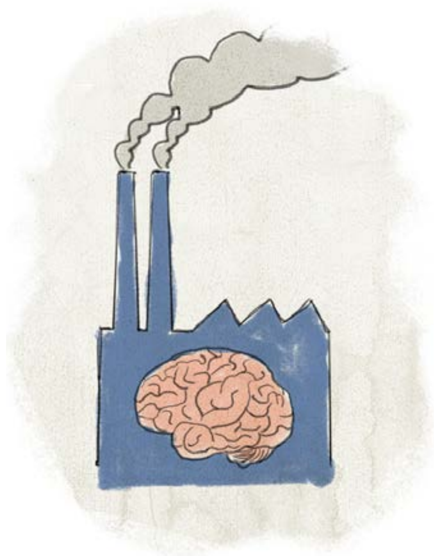
TABLE 1: **THE FEATURES OF EACH DATA MODE**

	Data mode			
	BUSINESS INTELLIGENCE	**DATA-INTENSIVE SCIENCE**	**WEB ANALYTICS**	**BIG DATA**
Data volume	Medium	Large	Medium/Large	Large
Data variety	Low – mostly structured	Low – mostly structured	Semi-structured	High – often unstructured
Data velocity	Slow (batch processing)	Slow	Fast	Fast
Data outputs	Dashboards and reports	New products	Dashboards, reports and experiments	Automated systems for decision-making
Data talent involved	BI and analytics professionals	Specialised academic disciplines e.g. computational biologists	Digital marketers	Data scientists

Business Intelligence mode

Business Intelligence companies tend to rely heavily on structured data about financial and operational performance and customer transactions, accessed through relational databases like SAS, and analysed 'in batch' in data warehouses and data marts.[14]

The primary outputs of Business Intelligence are descriptive, and include reports and dashboards tracking Key Performance Indicators (KPIs) to support decision-making by managers.[15]



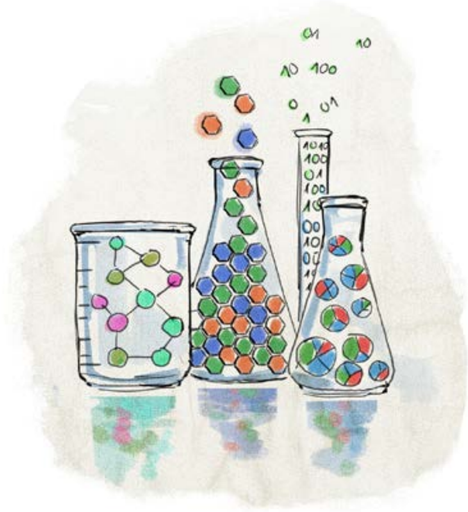
"I think the key point is that analytics and appropriate dashboards and self-service Business Intelligence is just much more widely available, so many more people in the company can access it. Now they are making decisions in a much more informed way, using data and fact. It's definitely helping them make informed decisions because they can refer to what's happening with this particular product, or what's happening with our customer base, the demographics of whoever, and whatever."

(Large Financial Services Company, FS3).[16]

Companies in this data mode generally recruit from a well-defined set of disciplines into widely recognised occupations such as Business Intelligence and Analytics roles.

Data–Intensive Science mode

Science–based industries like pharmaceuticals often have to work with large data sets such as genetic sequences and randomised clinical trial data. This data is generally structured:[17]



> “What’s different about the Pharmaceutical industry is a lot of the data that we collect to make decisions; it’s not data that is collected for any other purpose - it’s very targeted. We decide what data we’re going to collect, run the trial to get the data, make sure it’s validated properly, and then use that data to make a decision.”
> **(Small Pharmaceuticals Company, P7)**

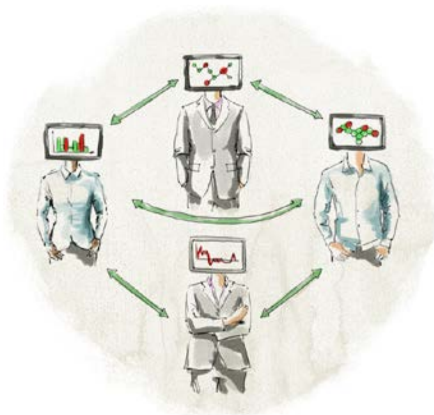
Data insights are used throughout the Research and Development process, including in drug discovery and in clinical trials. Although traditionally analysis has been rooted in the experimental method (collecting data to test specific hypotheses), access to more data is leading to a ‘Fourth Paradigm’ for scientific discovery where pattern recognition plays a stronger role.[18]

> “The questions that come back are much more complicated than what they were. So, someone’s been taking a drug for ten years, experiences an event. Is it the drug, is it complication, is it the condition, how do you answer that in a scope of having access to real world patient records?”
> **(Large Pharmaceuticals Company, P2)**

Data–Intensive Science businesses primarily recruit from well–established academic disciplines like epidemiology and computational biology.

Web Analytics mode

Companies in the Web Analytics data mode collect and analyse data from website users to track web and ad campaign performance, often in real time. Some of this data – such as event streams and server logs – is semi–structured, and can require substantial processing before analysis.



The analysis is generally descriptive, and often involves tracking high–frequency web metrics, and segmenting customers. ‘A/B tests’ are used to test the popularity of online features and content, and of promotional messages.[19] The results are used to redesign websites, improve conversions and increase the return on marketing investments.

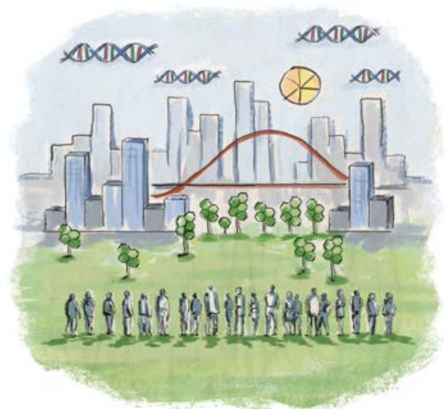
> Data–driven design is at its most powerful when it is responsive, when you can iterate fast. What that means is that, in order to run a test, you have to get the artist and the coder but also get the back-end programmer to rewrite the analytics database in order to track the data you need.
>
> **(Small Creative Media Company, CM8)**

Web Analytics companies employ digital marketers who use solutions like Google Analytics, Adobe Digital Marketing Suite and Webtrends.

Big Data mode

Recent advances in hardware, software and cloud services have made it possible for businesses to access and process massive volumes of data. This data often comes from the web, in unstructured or semi–structured formats, like text, audio and video content, web logs or network data – and at a fast velocity.

> Big data/data science refers to the arrival of new sources of data and a data infrastructure to make things scalable. There are more tools to work with data.
>
> **(Large Pharmaceuticals Company, P4)**



This has created substantial opportunities for innovation, linked to:

- Revelation: Data about users' online behaviours helps firms understand their preferences, as when a query in a search engine tells advertisers something about the interests of the user. This data can be mined for new and valuable insights.[20]
- Automation: Patterns in the data can be used to predict behaviours and outcomes, and deliver personalised services at scale and in real time.
- Learning: Machine–learning methods make it possible to improve the performance of big data–driven products and services dynamically, as they 'learn from their mistakes.'[21]

These opportunities are often best exploited by building new products and services e.g., search, and implementing new features into existing services e.g., recommendation systems in e–commerce sites, driven by algorithms based on these analytical insights.

Enter the data scientist

Innovating with big data requires new skills: for example, messy data has to be cleaned before it can be reliably analysed. Visualisation techniques beyond familiar charts are needed to discover and communicate patterns in complex data sets.[22] The insights from analysis are often implemented in data products.[23]

In the late 2000s, Silicon Valley companies working with big data like LinkedIn and Facebook started referring to workers with these skills as 'data scientists'.[24] The hybrid nature of their skill set is also picked up in our interviews:

> “A Data Scientist is an analyst or data analyst who also writes code – who also does code transformation of data. While people who just do data analysis, that’s a business analyst and we have a specific role for business analysts in our organisation. I also separate that narrow view from a Data Engineer who does not do analysis but just does engineering or data engineering. So it’s a relatively easy Venn diagram.”
> **(Medium Creative Media Company, CM4)**

> “Two previously separate areas have come together, programming and software engineering, and then statistical analysis and the whole mathematical side. I come from the engineering side. We don’t have any people coming through aside from those who have done it on the job who have those two blended. Those who would have strong mathematical abilities but also can program and understand the realities of running large–scale production systems.
>
> On the flipside, there are those who have software engineering and programming ability but don’t have necessarily the statistical analysis and modelling background, so understanding clustering etc. that we look at. So bringing these two strands together is currently the problem we’re tackling. Bringing these two streams together is the key in data science/analytics. It’s an approach used to understand and approach data and form hypotheses to be tested with data and then be able to build systems out of that.”
> **(Large Creative Media Company, CM9)**

In Box 2, we look at the relationship between data science and statistics.

The tools (and traits) of the trade

The data science field is evolving, and so is its toolkit, which mostly consists of open source (free to use and modify) resources including the Hadoop framework for parallel data processing, Cassandra (a data store optimised for large data sets), Pig (a language to query unstructured databases) and R (an application for statistical programming).[25]

Box 2: What is the relationship between data science and statistics?

There is a long-running debate about the relationship between statistics and data science amongst practitioners and educators (Shalizi, 2012). We also explored this issue with our interviewees. In general, they agree that statistics is a core component of the data scientist skill set – the ‘science part of data science’ (Wladawsky-Berger, 2014). As one of our interviewees put it:

> “Some data scientists I meet, they’re just data engineers, they’ve never done a regression analysis . . . so that’s not a data scientist in my book (…) A data scientist has to be at least partially working with statistics and knowing the math

behind statistics which a lot of data engineers are not that familiar with – or not to that degree. ”
(Medium Creative Media Company, CM4)

Having said this, some of our respondents highlight important differences between the way things are done in statistics and in data science:

“ The difference between statisticians and data scientists is that they approach problems/think differently: statisticians develop algorithms to solve a problem; data scientists scale those algorithms to serve thousands of users. ”
(Large ICT Company, ICT6)

“ Traditional statisticians look at a reservoir, and data scientists look at a river flow of data (including third party data sets) to identify patterns there that can trigger real-time actions. ”
(Large Creative Media Company, CM6)

This distinction reflects the computer science and engineering aspects of data science: data science is not just about writing research reports and papers, but also building new tools and products (Dhar, 2014), which requires coding skills. This problem-oriented feature of data science makes it multidisciplinary and pragmatic, and very open to borrowing techniques originating outside of statistics, such as machine learning algorithms from computing, graph theory from maths, and text-mining techniques from linguistics. (Breiman, 2001, Cleveland, 2001, Varian, 2012). Of course, some self-defined statisticians do work in this way, but it is perhaps less part of their formal training and approach.

Another distinction sometimes made is that data scientists tend to be more focused on big data – that is, large unstructured data sets. Some statisticians argue, however, that their strength is in the analysis of data sets (large or small), and that some data science practitioners may be good at collecting and storing data, but that their analytical skills and knowing what the data can actually tell them may be weaker.

It is clear that linguistic debates in this area will continue for some time. More prosaically we have heard of statisticians who are changing their job title to ‘data scientist’ as a way of gaining recognition and a financial premium in the job marketplace.

Going beyond specific tools, we hear that data scientists have a ‘creative’ and passionate attitude to data that combines the rigour and curiosity of the academic researcher and the pragmatism/resourcefulness of the hacker.[26] Like academics and software developers, data scientists often participate in innovation communities – this includes contributing to open source projects and expert websites like Stack Overflow, and getting involved in meet-ups and hack-days to share experiences and skills. This helps them showcase skills and learn about the latest developments in the field:

“ I find it more useful to send people to meet–ups because lots of meet-ups do ‘how-to’ sessions and walk you through a process of working with a tool for an hour or two. Or meet-ups for a weekend of hacking data where you can go and pick up some new skill sets for free – just by going to the meet-ups. Someone will go to a conference and pick up a new skill set and mentor the rest of the team together. ”
(Medium Creative Media Company, CM4)

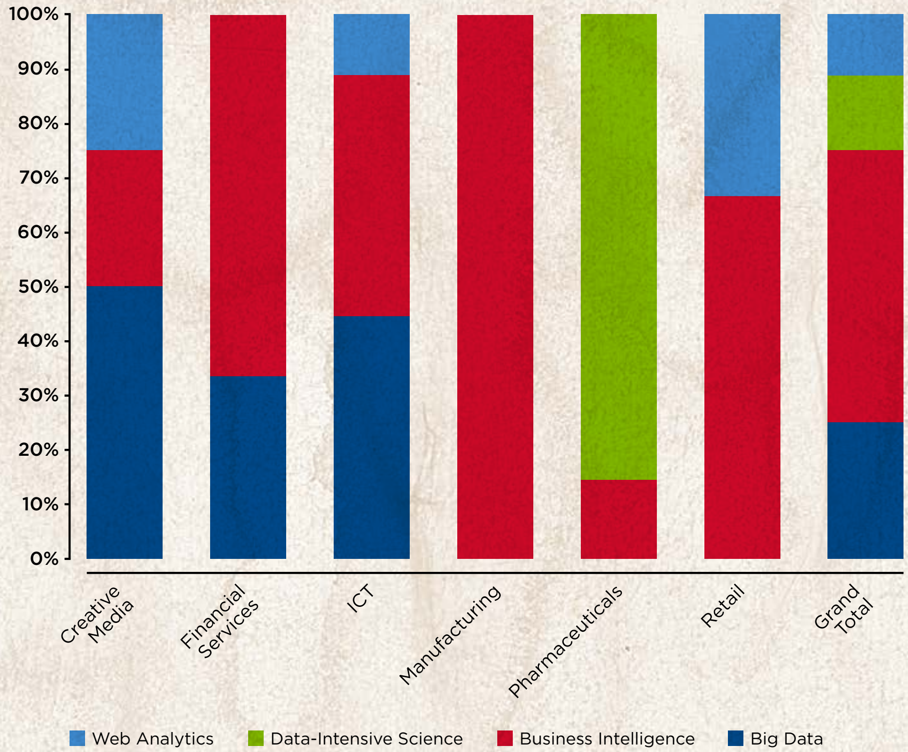
II. Big data convergence: how the data landscape is evolving

Big Data companies are in the minority, but perhaps not for long?

As Figure 1 shows, businesses operating in the Big Data mode make up around a quarter of the organisations we interviewed.[27] They are primarily Creative Media, ICT and Financial Services companies providing online content and services – for example, social games, web advertising and online banking.

Most Pharmaceuticals companies we interviewed specialise in Data–Intensive Science, and all our Manufacturers follow a Business Intelligence approach. Retailers with large 'bricks and mortar' operations also use Business Intelligence, while e–commerce companies employ Web Analytics.[28]

FIGURE 1: **DATA MODES BY SECTOR IN OUR SAMPLE**



This picture is not static though. Although there remain uncertainties – and in some cases scepticism – about the hype surrounding Big Data (as one of our interviewees puts it, it can feel like 'a solution looking for a problem'), the overwhelming majority of the companies we interviewed are taking steps, albeit to varying degrees, in a Big Data direction. As one of our interviewees put it:

> "Big data is overhyped. Those who are sceptical about it are wise, but too much scepticism would be foolish."
> **(Large Financial Services Company, FS9)**

The change we see is being driven by supply and demand-side factors.

On the supply side, more and more data sets – including open government data and unstructured web data – are becoming available for analysis, and new technologies and cloud services are lowering the costs of data storage, management and analysis. Our respondents are also bringing together a variety of data sets previously kept in siloes with the goal of gaining a 'single customer view':

> “Now we record every item of our stock, all the way through its life, every single day, until it's bucketed and disappeared, so that's 12 billion records a year. Even two years ago, we couldn't have imagined doing that.”
> **(Large Retail Company, R5)**

> “There are more sources out there and a lot more desire to use those sources. (What triggered the idea for investing in data analytics as a business tool) was understanding that there were new sources of information and that our business is changing and we have to show additional value for our products, and having information to describe the value becomes more and more imperative. We also understand that information is critical in finding new products so there is that emphasis and finally everybody is doing it.”
> **(Large Pharmaceuticals Company, P5)**

On the demand side, there is an increasing recognition of the commercial value of large-scale data analysis, in some cases inspired by the activities of Big Data companies like Google and Amazon, in others by changes in customer expectations (for example, the growing demand for personalised services).

> “I think the most important thing that is going to happen in the next years is that patients will have much more of a say as to what is happening to their healthcare and they will drive information flow. We talk about the healthcare ecosystem evolving and at the core will be patients and they will demand information about themselves and ways of how they might benefit and about their ability to almost decide the best path forward for them in healthcare.”
> **(Large Pharmaceuticals Company, P3)**

The result is much experimentation. Pharmaceuticals companies are using unstructured data from internet forums to evaluate drug safety, manufacturers are exploiting machine sensor data to speed up product development, and retailers are mashing up various data sets to decide where to locate their stores. Companies operating in the Business Intelligence data mode are applying sophisticated data-mining algorithms to move from description to prediction. As a consequence of all this experimentation, the demand for new types of data talent (including data scientists) is growing rapidly in all sectors.

3. The labour market for data talent

I. What makes the perfect data analyst?

One of the most striking findings from our interviews is the remarkable similarity in the profile of the data talent that companies are looking for. Regardless of which data mode best characterises the business. Many of the companies that we interviewed are demanding people with the mix of skills used to define data scientists – including analytical and programming skills, and business knowledge (see Table 2 for the profile of the perfect data analyst). We think that this is being driven by the increasing awareness of the opportunities that big data present for innovation in all sectors.

TABLE 2: **THE PROFILE OF A PERFECT DATA ANALYST**

	Description
Core skills	• **Analytical:** strong mathematics and statistics foundation. • **Technical:** programming and database skills.
Domain and business knowledge	• **Knowledge of the sector:** understanding data sources and real-world situations and processes behind the data. • **Awareness of business goals and processes:** Knowing the business questions that matter, and how data fits with the organisation.
Soft skills	• **Storytelling:** Ability to transform analytical insights into actionable – and compelling – business recommendations. • **Team-working:** Enjoy working with people from different disciplines.
Competencies	• **Analytical mindset:** Being able to re-formulate complex questions as analytical tasks. • **Creativity:** Knack for generating unexpected solutions to problems, and exploring data from different angles. • **Curiosity:** Wanting to understand how the world works.

The core skills

All our interviewees agree that a strong analytical foundation is essential for a data analyst – the clear majority of our interviewees prefer to recruit people from quantitative disciplines like statistics, applied statistics, the physical sciences, computer science and engineering. They also look for candidates with strong computing skills including computer programming, and database design and operation.

> “The key skills are deep understanding and data basing skills, high performance computational science skills, the analytics aspects, the mathematics and statistics and AI, that sort of thing, and then the business acumen.”
> **(Large Pharmaceuticals Company, P5)**

These skills are widely seen as transferable across sectors:[29]

> “I found these skills are incredibly transferable and we recruit analytics staff from other sectors e.g. manufacturing, finance and retail as they do the same thing we do and have large data sources. For example, someone in finance uses data as predictive models of how someone is likely to default on their mortgage. We are doing predictive models based on patient attributes – how likely they will be hospitalised in the next six months. The topic is very different, but the techniques are very similar.”
> **(Large Pharmaceuticals Company, P5)**

Regarding skills levels, PhDs are particularly valued insofar as they prepare individuals for solving problems conceptually and, in some cases, working with large data sets.[30]

> “When we talk about data and data sciences, in some ways you are talking about research and research methods; quantitative and qualitative analysis, hypotheses testing, and that fundamental and deep understanding of research methods using data (collecting, correlation, causality) you tend to only find that people who’ve done PhDs, less so MAs.”
> **(Small Creative Media Company, CM5)**

Domain and business knowledge

Domain knowledge is also considered important by our interviewees. In some cases this means strong sector-specific expertise – understanding the theories that explain relationships in the data, and knowing data sources and their limitations:

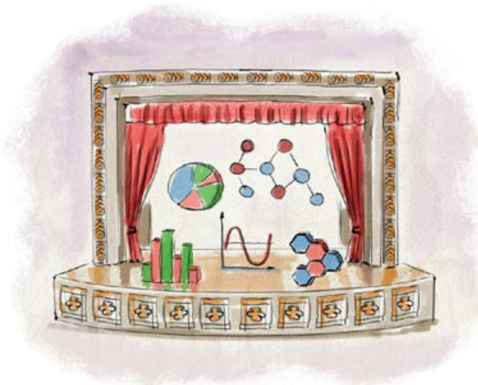
> “I think you would normally be looking for a science background, typically a biological science background, and that just really matches with the type of data that we’re collecting.”
> **(Small Pharmaceuticals Company, P7)**

Businesses also say they want applicants who have a general knowledge of business processes and goals, and who understand that data analysis is not an end in itself, but a tool to achieve business outcomes. This business knowledge allows data analysts to identify those questions that are important for their organisation:

> “Again, I think the danger is the data science tag is bringing in a lot of people where the coding and the science bit is more important; they are fascinated about how to do the analysis rather than what the analysis is actually going to deliver to the business. (…) You talk to some people who are data scientists and they absolutely get it, they understand that anyway. You get others where you get very arcane debates about coding and R and Python and God only knows what else.”
> **(Large Retail Company, R3)**

Soft skills

Many of our interviewees say they look for data analysts who can use data to tell compelling stories:[31]



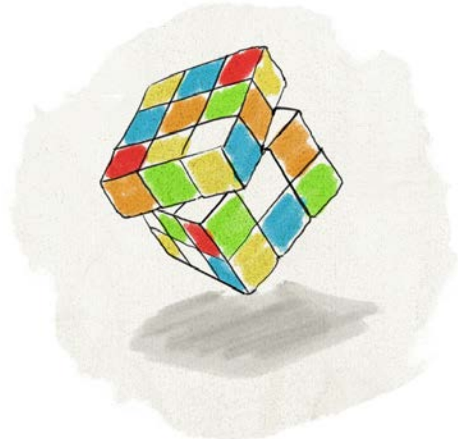
> “There is such a massive need to balance the technical side of things, the ability to handle data and the ability to translate it into something that means something in a business context. That’s the bit that is most valuable to us. You can turn a few million lines of data into a table, but unless you can read that table and talk about what the implications are in the context of whatever question is being asked or whatever changes are being proposed, then it’s useless information.”
> **(Large Financial services company FS4)**

The ability to work with people from other disciplines is also important. This does not just refer to other analysts, but also to people in different parts of the business that might be less data savvy – including managers and domain experts in other departments:

> "The predictive models that we use need to draw on the domain/subject expertise of others in the organisation. However, it can be difficult to build communication between analytics experts and domain experts because of differences in language, and lack of trust."
> **(Large Public Sector Organisation, O2)**

Competencies

The competencies or attitudes that our interviewees look for in new data talent include an analytical mind-set (the ability to take a complex issue and transform it into a set of analytical tasks that can be tackled using data), curiosity (a passion for understanding how the world works, and for finding interesting patterns in data) and creativity (the ability to combine ideas from different domains in order to find unexpected solutions to problems):



> "Human Resources here had kittens with me because I said I wanted to leave a half-repaired radio or a Meccano set not completed, or some sort of Lego half built in the waiting room, and just wait and employ the guy who tried to work out how the Meccano should fit together while they were waiting (for their job interview)."
> **(Large Retail Company, R3)**

II. Desperately seeking data talent

What is the experience of the companies we interviewed in the market for data talent in the UK? Are they finding people with the skills that we have described? The answer, consistent with previous studies, is a resounding no. The overwhelming majority – as much as four in five of our interviewees – reported difficulties recruiting data talent in the UK.[32]

The available data talent is not up to it…

The most common explanation for skills shortages is that candidates lack technical skills and hands-on experience working with data. The companies we interviewed are caught between a rock and a hard place: recruiting experienced talent is extremely costly; inexperienced recruits are cheaper, but they can take months to start contributing to the business:

> "The skill set in-house inside the country is quite limited. To an extent, with university degrees, they're not… practical, tied to what the business requires…. If you're taking on a graduate, then you need to train them for about a year, to get the results you want. It's fine to train people from university but not to re-teach them."
> **(Large Manufacturing Company, M4)**

A small number of interviewees are worried about the average quality of UK graduates compared with those from other countries like China or India, as well as from Eastern Europe:

> "At (previous company), we used to take a lot of Indians and Chinese. The quality of the education in order to deliver this skill set was in advance of your average UK. If you're only going to recruit from Oxford, Cambridge, Imperial, then you're probably alright, but not everybody wants to come work for a start-up. If you go out of the top three or four universities, you could probably get better abroad."
>
> **(Medium-sized Financial Services Company, F7)**

The supply of data talent is insufficient…

Other interviewees say that the volume of data talent is insufficient to match demand, which again makes it expensive to recruit. Pharmaceutical companies in particular feel that the labour market for certain data disciplines is thinner in the UK than in other countries:

> "For the past few years the UK is not a great place to find people who are very numerically trained. I say that quite generally. I think it's a good place to find general statisticians – actually, when you get down to epidemiologists, or informaticians (…) that's always been a big gap to have that high–end people compared to the US."
>
> **(Large Pharmaceuticals Company, P2)**

This problem appears to be more severe outside London, where companies struggle to attract data talent. One of them goes as far as saying that they are considering opening an office in London primarily to access data talent. At the same time, some of our London-based interviewees are concerned about 'overheating' in the data labour market because of growing demand from companies in the Tech City digital cluster in East London.

It is hard to find people with the right mix of skills…

Some companies we interviewed say that the problem is not finding discipline specialists, but people with a hybrid skill set. Most candidates may be good analysts, but poor coders, for example (or the other way around):

> "Right now, we have a situation where we have 'kids' under 30 who are very, very talented in programming and we try to get them to use things like R and they can implement it but they don't know what it means."
>
> **(Small Creative Media Company, CM5)**

In other instances, data analysts who are technically proficient lack the domain expertise and business awareness to thrive in industry:

> "You have to have somebody with a very strong technical background, but you usually find the ones with the very strong technical background are not the ones adding the most value, they are the ones crunching the numbers and they are

not adding the insight to the numbers. Definitely everybody that I've interviewed in the past two years for these specific roles …you just get two types of people, actually you get three types. One that's very, very technical… would be great if you could lock him in a room and he doesn't have to do anything else. You get another one that if he didn't have to do any technical work he'd probably be decent. And you get the rare type, which is the combination, which is very, very hard to find. ”

(Large Financial Services Company, FS1)

Businesses lack in-house expertise to recruit data talent…

Internal confusion about what data analysis involves also creates challenges for effective recruitment, particularly for companies experimenting with new types of data and analytical techniques.

For example, some of our interviewees say they struggle to convince their HR departments that data analysis roles are different from IT roles, and end up having to sift through piles of irrelevant applications for those jobs. Finding relevant candidates through recruiting agencies is also challenging (and expensive).

Managers who are not themselves data specialists are unable to assess candidate skills (this problem is intensified by analysts getting on the 'data bandwagon' and padding their CVs with data-related buzzwords even though it is unclear what their actual skills are). Insufficient appreciation of how data can create value inside a company also limits the salaries on offer for data analysts:

“ There's not enough understanding of it at the senior executive level in the company, I don't think I'm speaking out of turn here. I'm not sure there is a full enough appreciation of it, though they might challenge that. Just education and proving to them in practice. I think they have to see things, rather than just explaining theory, they need to see actual impacts on the day–to–day operations from using data analytics. ”

(Large Financial Services Company, FS2)

III. Where to now? The future evolution of the market for data talent

Over time, there are some forces that could alleviate the skills shortages we have just described. This includes better analytics tools that reduce the need to hire data analysts, changes in the educational offer in response to business needs, and an improved understanding of data analyst roles and functions inside business. Will these changes be sufficient to remove the recruitment challenges reported by our interviewees?

Some expect skills shortages to become less severe

Universities are starting to combine analysis and computer programming in new degrees, and this should make it easier to recruit graduates with the right skill set:

“Some academic programmes are trying to mix computer science with statistics and business and I am starting to see some master’s programmes in analytics so there is promise.”
(Large Pharmaceuticals Company, P5)

“Five years ago academia didn’t really want to talk to us, thinking it was a smutty industry and they didn’t want to deal with us. I think they’re beginning to change. They are now keen to talk; it’s just occasionally they want different things than we do from the process, but it’s getting there.”
(Large Retail Company, R3)

Others expect that new tools will automate and commoditise many current data analysis functions:

“We have a new product offering (product name) which is designed to address the skill shortage we see around data scientists and to allow customers to get started with small to large data sets more quickly by providing an easier interface where the tool will automatically go and run a large number of (models) and return to the user findings ranked by interestingness.”
(Large ICT Company, ICT7)

Some of our interviewees go as far as predicting a future where a group of elite data workers develop easy–to–use tools that are applied by a mass of data savvy users, with analysts who are ‘in the middle’ all but disappearing from the workforce.

Others think the market for data talent will only heat up further

There are other forces at play, however, that, according to some of our interviewees, will intensify the demand for data talent. Most obviously, there are the increasing volumes of data that can be analysed, and increased competition for talent from different sectors:

“Skills shortages will get worse. I think in the future it will become an issue. I can think about the volumes of data that we have to deal with compared to what it was years ago,(…) and the ability to do that and the skills required to do that will just increase.”
(Medium–sized Manufacturing Company, M6)

“More competition between sectors – we’re just competing against lots of different industry sectors in (Name of company); it just makes it difficult, we have to compete with banks, professional services as well as competing with our own industry, it slows it down and increases the amount of work for people.”
(Large Creative Media Company, CM2)

There are also doubts in the minds of some respondents about whether the UK education system is able to respond fast enough to the demands of industry, and deliver the data talent that is needed:

> “I can see business schools are starting to create some Information Systems courses. While they tended to be on the Engineering side, the Computer side, it’s shifting slowly, but I don’t think there’s enough promotion. It will be totally irrelevant if (…) we don’t get there in a very short period.”
>
> **(Large Manufacturing Company, M4)**

> “I expect that is probably going to get worse as time passes because it’s not likely that the market signals that there’s a shortage will get through to young kids when they’re considering what courses to do at university, so I expect it will just get worse. I think it’s for anyone who deals with data; it’s not exclusive to marketing or banking or insurance.”
>
> **(Large Financial Services Company, FS4)**

One of our interviewees pointed out that more and better tools for data analysis might in fact increase overall demand for data analysts by allowing smaller businesses to set up their own data teams:

> “(Companies that) would’ve just outsourced the analytics to an Experian or Equifax, because they couldn’t have afforded the entry cost to do it themselves, and it wasn’t worth doing in-house. It’s now a lot easier to build a team of two or three people so there’s a lot of demand increasing quite rapidly.”
>
> **(Large Retail Company, R3)**

There is also some scepticism about the extent to which tools will be able to truly replace data analysts:

> “Most companies are developing technology solutions to talent problems. Data analyses are made available to people with lowest common denominator skills, people who are not able to interrogate the data intelligently. American companies are proficient at producing user-friendly Graphical User Interfaces that simplify and automate, but it isn’t clear who is checking the work. This creates situations where ‘big data’ can create (and compound) big mistakes.”
>
> **(Large Creative Media Company, CM6)**

It is also unlikely that off–the–shelf tools will satisfy the skills needs of companies at the very cutting edge of their sectors. This is because these companies’ capabilities to create value from data change rapidly as new data comes online, whereas automation is based on past best practices and methods. As a consequence, the competition for high–end talent among those companies can be expected to remain fierce.

Data offshoring

Finally, some companies say that they are considering offshoring, or have already offshored, UK data analysis capability to locations where talent is easier or cheaper to recruit, such as India and China:

> "I think in industry we'll see a lot more offshoring, potentially to deal with the skills shortage. Looking to China or India or other low cost base countries, the staff would probably be based over there, the cost effect is stronger if they reside in their home countries. I can see that happening in terms of increasing internationalisation of skills and people. You could then tap into a wider base of analysts."
>
> **(Large Retail Company, R2)**

Some express concerns about the long–term implications of this offshoring for the UK skills base and their companies' ability to deal with data.

> "But for the long term I see it as a risk to the UK and US industries in this field because the people onshore aren't going to have the skill sets to fill those roles. I think it's less about skills and more about experience, so they're not going to have the experience of working on and managing complex data capture."
>
> **(Small Pharmaceuticals Company, P7)**

5. Implications for educators and policy: improving the supply of data talent in the UK

Although making any predictions about the future evolution of the market for data talent is hazardous, what appears clear is that UK-based businesses will, in the near term at least, continue to suffer from recruitment difficulties.

This should be a concern for policymakers, because it hinders UK companies' ability to compete in data-driven markets, where there are often strong first-mover advantages (think of social networking or e-commerce, for example). It might also lead UK businesses to offshore their data capabilities in a way that erodes the UK's analytical skill base, and also reduce the attractiveness of the UK as a location for global companies looking to invest. In the short term, there are several areas we detail where action can be taken to improve access to data analysis skills for UK businesses caught in the midst of what seems to be a data talent crunch.

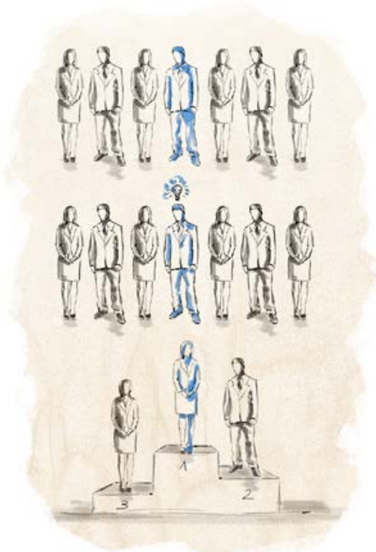
Our findings also have implications for higher education: many of the companies we interviewed complained about problems with the data talent graduating from UK universities. Some are also concerned about the quality of analytical training in schools (compounded by negative perceptions in the eyes of young people of data analysis as a career), and the longer term implications this has for the supply of UK data talent.

In the rest of this section, we identify seven areas for action, drawing on the suggestions from the companies we interviewed.

Skills

1.Develop the skills of the UK data workforce

Professional bodies, such as the Royal Statistical Society, and the Sector Skills Councils should consider ways to develop the skills of established professions like statisticians, computer scientists and social scientists to fill the skills gaps in business. As we mentioned above, in some instances candidates already have one of the essential ingredients (e.g. strong analytical or coding skills, or domain knowledge) but are missing the other. Universities could play an active role through Continuous Professional Development (CPD) programmes and short training courses to build pools of talent with the mix of skills that business need.



This effort also has a local economic development angle, and not just because companies outside of London appear to be facing more of a struggle when they try to recruit data talent. Research in the US suggests that there are local knowledge spillovers from company investment in big data skills – that is, companies benefit from investments in data talent by others in their vicinity.[33] In England, Local Enterprise Partnerships (LEPs) should consider how they can encourage these spillovers, and kick-start virtuous cycles of data skills development, by bringing together local universities with industry to improve the supply of data talent. They should also explore if there is scope for targeted incentives to encourage business to train their data workforce, perhaps with support from existing programmes like the Department for Business, Innovation and Skills' Growth and Innovation Fund and Regional Growth Fund.

2. Build up the data analyst profession

A number of our interviewees suggested that intermediary organisations such as industry associations and professional bodies have a critical role to play in developing data communities of practice and establishing training and qualification pathways:

> “Maybe an industry body would be an appropriate way of bringing people together, the organisations so they can get a wider view of what's needed and then can push institutions to promote these courses.”
>
> **(Large Retail Company, R2)**

An example of one such body in the UK is the recently established Society of Data Miners (SocDM), which aims to 'increase the benefits of data mining to society by raising awareness and bringing the analytics community together as a true profession.'[34] More established bodies such as the British Computing Society, the Royal Statistical Society and the Operational Research Society all cross over into this emerging field. It remains to be seen whether a single industry body arises in this area, if some collaboration between existing bodies emerges, or if the field remains relatively fractured.

In addition to developing certification standards, liaising with education and raising awareness, stronger professional bodies can also act as a forum to showcase good practices for the management of data talent, and improve knowledge-sharing between industries.

> “I think there is this potential that we start to get this cross-industry discipline talking going on to see a bit of interchange from people moving from one sector to another. I think that has got the chance to bring skills bridges that we would never have seen before.”
>
> **(Large Pharmaceuticals Company, P2)**

3. Ensure access to overseas talent

Given the acute nature of the skills shortages, policymakers should consider how they can attract skilled immigrant labour (including international students in courses related to advanced data analysis). Some of the companies we interviewed were concerned about a further tightening of immigration policy:

> "It's all about skill shortage and I think the biggest threat is probably immigration. We already have problems because we are cut off from American talent and Indian and Chinese talent. If we are potentially cut off from European talent I don't see how London as a tech sector can survive. We need that talent to move forward."
>
> **(Large Pharmaceuticals Company, P3)**

Some options include improving the UK employment prospects for foreign graduates after they complete advanced data analysis courses at UK universities. Recent policy changes have acted in the opposite direction; in particular, the closure of the Tier 1 Post Study Work (PSW) route, which has reduced entitlement of international students to work in the UK, and more stringent university accreditation procedures. The UK government needs to project a more positive image of the country to overseas data talent if it is fortunate enough that the talent wants to work in the UK. The government should also consider extending the Tier 2 Shortage Occupation List to incorporate data scientists.

Education

4. Establish closer links between employers and universities in the data area

Many of the businesses we interviewed complained that university curricula remain too theoretical, and fail to reproduce key features of 'real world' data analysis work (not least having to work with big and messy data sets).

The result is that students graduate lacking important skills for industry. Sectors Skills Councils should play an active role in communicating to universities the skills needs of the industries they represent, and universities should consider how they can adapt their curricula in order to address those needs (this includes exploring options to combine traditional lecturing with the wealth of high-quality content in Massively Open Online Courses (MOOCs), as well as hands-on work with free tools and open data sets).

New fora bringing together universities and data-intensive businesses should be explored (perhaps under the auspices of the National Centre for Universities and Business (NCUB)).

The recently founded US National Consortium for Data Science (NCDS) is an interesting model, aiming to 'help the US take advantage of the ever increasing flow of digital data in ways that result in new jobs and industries, new advances in healthcare, transformative discoveries in science, and competitive advantages for US industry.'[35] It sets out to achieve this through a data observatory (providing a shared infrastructure where members can access big data sets for analysis), a data laboratory (a test-bed for experimentation with big data technologies) and a data fellows programme to train 'tomorrow's data science leaders.'[36]

In principle, given its focus on unlocking value across the 'data value chain', the Connected and Digital Economy Catapult (CDEC) could play a similar role to NCDS (although its current focus is on R&D and commercialisation rather than skills development and training). The Data Science Institute at Imperial College is also an interesting example of a 'single point of contact' for companies interested in big data that other universities might want to imitate.

5. Improve the supply of data talent with hybrid skill sets from education

Our research strongly supports the idea that UK universities are failing to produce graduates with the skills mix that data-driven businesses want. Perhaps this should not be a big surprise: there have long been concerns about how funding and organisational factors create hurdles to interdisciplinary teaching and research in UK universities.[37] While, say, the US major/minor system allows students to choose two fields of specialisation when they do a degree, UK students typically specialise in a single field. Anecdotally, one of our interviewees mentioned that PhD programmes in the US are in general broader and less specialised than in the UK too.

Research Councils and universities need to think hard about how they can create incentives for the sorts of interdisciplinary teaching and research that will produce tomorrow's data talent. One of our interviewees used Stanford University to illustrate how an institution can (literally) build structures for collaboration and knowledge-sharing across disciplines.

> "In Stanford, Computer Science, Information Science and Statistics buildings are located in a triangle 100 meters away from each other. This is not an accident. Proximity means that there is cross-pollination (and cross-listing) of faculty, and also a flow of students between departments."
>
> **(Large Company Financial Services, FS9)**

There are some initiatives in place to encourage disciplinary linkages in the UK. The Network on Computational Statistics and Machine Learning (NCSML) funded by the Engineering and Physical Sciences Research Council (EPSRC) brings together experts from the statistics and Artificial Intelligence fields in workshops, seminars and conferences.[38]

Meanwhile, the Nuffield Foundation's Q-Step programme aims to boost the quantitative analysis skills of students in other disciplines (in this case, the social sciences). This is being done through 15 Q-Steps centres which are developing innovative courses, work placements and pathways to 'promote a step-change in quantitative social science training in the UK.'[39] There is no reason why a similar initiative could not be put in place to boost the programming skills of statistics graduates – or the analytical skills of computer science ones.

6. Strengthen the UK education system to improve numeracy and data handling across the board

The long-term supply of data talent for UK businesses depends, of course, on what happens in schools. Ensuring that they equip young people with numeracy and data-handling skills is critical, as these are the basic building blocks for future data analysts, as they are for other STEM occupations. The UK is however at the bottom of the league table when it comes to quantitative skills.[40] The introduction of new 'Core Mathematics' qualifications is a welcome step in the right direction, generating options for mathematics study post-16 for those who do not take A-level mathematics. We need to ensure that data handling, problem solving and statistics are core elements of mathematics education.[41] These skills are not always easy to assess through exams, and so thinking also needs to be done about method of assessment in these areas.

In addition to ensuring that the mathematics school curriculum is fit for purpose, and that educators have the right skills to teach it, schools, industry and parents should look for ways to harness the current proliferation of coding clubs and after-school tech activities to introduce young people to creative data analysis (potentially using the wealth of open data that is becoming available). Attempts to adapt education models like School of Data which have been successful in other settings to schools should be encouraged:[42]

> “One of the things that could happen is more focus in schools using those data sets for projects or within A-level syllabuses. There's a huge amount of data on Data.gov. Are 17 year olds pulling out APIS, ASBOs, traffic offences, and using that? That makes it interesting and often you're using the last quarter's data ... and pulling out quite interesting insights from it.”
> **(Medium-sized Financial Services Company, F7)**

7. Change long-term perceptions of data as boring, and data-related jobs as uncreative

Our research has confirmed that data analysis work involves a high degree of creativity, and is central to some of the most exciting and innovative businesses in the UK economy, from developing new games to discovering new drugs. Yet, in spite of this, our interviewees complained that there remains a perception that data is associated with dull jobs, and also that data analysis is only relevant to a small number of industries.[43]

> “I think there's a lack of awareness that there are careers in this field, I think that people that could be really good at this, when making career choices, won't necessarily know that these roles exist, and not just that they exist, but exist outside of Google as well. There will be people who love fashion, and love working with computers and data, and they probably have no idea that there's a way they could bring those two things together.”
> **(Large Financial Services Company, FS4)**

The companies we interviewed agree that in order to address this, businesses need to find ways to communicate the importance of data as a driver of innovation and growth across the UK economy. This is something that itself will require a great deal of creativity, but it is not impossible, as made clear by the success of the Next Gen Skills coalition in getting educators, parents and young people enthused about computer programming, a subject that many had previously considered dry and dull:

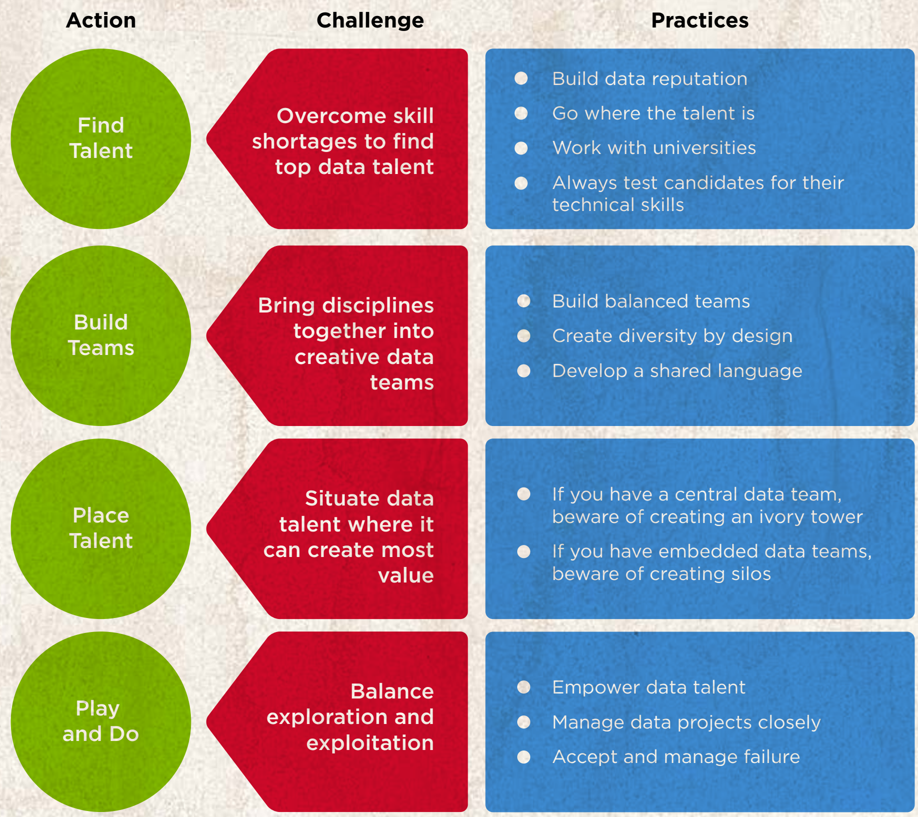
> “I think there is a responsibility for business to describe how data is really important. Pharmaceuticals is a good example; you'll get lots of things where you'll see pipettes and "This is how we create a drug." No, you do an unbelievable data analysis at the back–end. It's the data analysis which is the important bit; (the pipette) is the generic piece. How do you bring that alive and talk about that story and make it exciting?”
> **(Medium-sized Financial Services Company, F7)**

5. Implications for managers

What does our research mean for managers? It paints a challenging picture: while everyone agrees on the importance of data, it is not clear how to exploit it, and the talent that is essential for doing so is hard to come by. There are very few people out there ('unicorns') with the mix of skills that industry wants. The challenges are compounded by the fact that data analysts are creative individuals, doing innovative work – motivating such workers, and managing risky projects only adds to the difficulties.[44]

In this section, we describe how the leading-edge companies we interviewed are dealing with this situation, building their data capabilities and creating value from their data (Figure 2).

FIGURE 2 **PRACTICES TO BUILD DATA CAPABILITY**

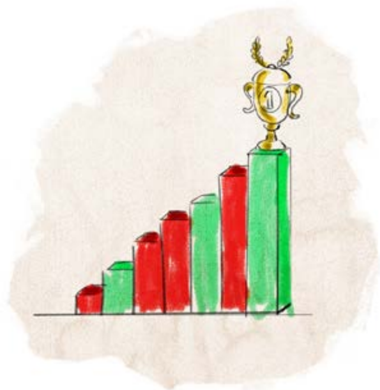


I. Finding talent

The companies we interviewed are adopting a variety of approaches to stay ahead of other employers in the race to recruit data talent. This includes:

Building a data reputation

One way of attracting data talent is by building a reputation for doing interesting work with data – in other words, by being seen as 'a place to be' for data talent.

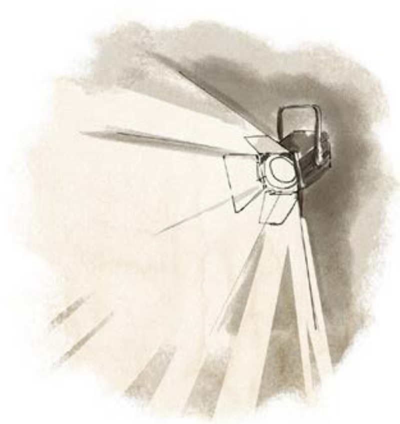


> "Most people who apply to LinkedIn, or Facebook or Google, they don't go to recruiters, or jobfairs or anything. They apply directly because they think what a company does is amazing. And that's in large part saying what you're doing in public – and the technologies you're doing and what you're using and the cool stuff you do with technology."
>
> **(Creative Media Medium-sized Company, CM4)**

This can be achieved by contributing to open source projects, publishing interesting (open) data sets and findings, and speaking at public events and conferences. Companies operating in less data-intensive sectors can turn this situation to their advantage by presenting themselves as 'greenfield sites' where ambitious data workers can make a difference, and enjoy the opportunity to work with new and interesting data.

Going where the talent is

As we said, data talent (especially data scientists) are often active participants in online communities such as Stack Overflow, LinkedIn groups, and competition sites like Kaggle. They also like to attend meet-ups, hack-days and conferences. Some of the companies that we interviewed frequent these places in the lookout for talent:



> "I always like (going into fantasy-land) to look for talent in places I know people who want to push themselves or who are ambitious might be, so I just recruited a guy from a hack-a-thon in Shoreditch. I will look at the people I see there and the top 50 (hackers) under 30 and lists like that in magazines and go after people."
>
> **(Medium-sized Financial Services Company, F7)**

Working with universities

Some of the larger companies we interviewed, particularly in the pharmaceuticals sector, have put in place specialist data graduate and work placement programmes. Others in industries not traditionally known for hiring data talent have started holding lectures at universities, and mentoring students to signal that they are in fact active in this area:

> "We did a lecture about big data and aspects of it. (It is) important for us to get our name out there as a place that would be interesting to work. Put ourselves in front of students and community building and building bridges into education, not necessarily via the traditional milk round open day, we try to go in at a grass root level, lectures and getting involved with projects."
>
> **(Large Creative Media Company, CM9)**

Always testing candidates for their technical skills

All of the companies we interviewed conduct in-depth testing of the technical skills of their applicants. Companies lacking the skills to do this internally are advised to work with external experts to design and evaluate those tests.

II. Building data teams

'Finding unicorns' with the right mix of skills is an arduous task. In response to this, the companies we interviewed are opting to build teams that bring together the variety of skills needed to create value from data. As Booz and Hamilton put it recently, 'data science is a team sport'.[45] In putting together these teams, the companies that we interviewed are:

Creating balanced teams

What is the right balance between specialisation and generalism? Specialisation can enable companies to operate at the cutting edge of a smaller area, but reduce their flexibility to resolve problems elsewhere. In general, we find that larger team sizes allow for more individual specialisation: large companies can afford to recruit deep experts (people working in 'niches inside niches'), while smaller companies tend to opt for generalists who can turn their hand to different tasks:

“This is in part because specialist skills can be underutilised. Many of the problems (our data scientists face) can be optimised along many dimensions of quality. You want a single individual to own a problem like that, instead of splitting it up between individuals. A generalist is preferable because he will be able to perform well across several dimensions of the problem, while a specialist will excel in a single dimension but will need help from others for everything else.”

(Small ICT Company, ICT1)

Creating diversity by design

Several companies we interviewed stressed the importance of building diverse teams. Diversity increases the probability that someone in a team will be able to deal with unanticipated problems. Diversity also generates opportunities for learning inside the team, which is an important motivator for data analysts.[46]

“The best teams I’ve ever seen are the ones that have a combination of unique individuals or across the team, to be able to understand what the data is showing, and the ability to feel and cross–reference that versus the real world (…)The teams that are most productive have the ability to mine the data for unusual insights, non–obvious insights, and be able to identify that in the real world, and then be able to take it to, “So what for the business?”.”

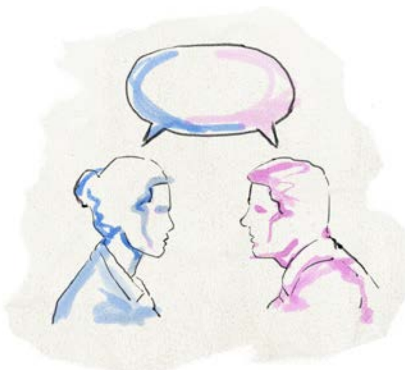
(Small Retail Company, R1)

Developing a shared language

Building a data team is not just about the people. It also involves developing a shared culture, language and standards, and putting in place systems to ensure that project lessons and outputs (including intermediate outputs, like analytical scripts) are preserved for future use.[47]

“I think productivity requires an ability to reuse information and not reinvent the wheel every time. To set certain standards about definitions so you only have one version of the truth and things like that… Re–use, and using what you’ve done before as a repository you can draw on for the future rather than it just being a throw away each thing done by itself, and it has no value in the future.”

(Large Financial Services Company, FS2)



III. Placing talent

Choosing where to locate data talent inside the business is a big question for our interviewees – and the answer depends on the company. In some cases, firms create a central data analysis unit serving the needs of other departments. In others, data analysts are embedded across the organisation.[48] Each approach has its pros and cons.

If you have a central data team, beware of becoming an ivory tower

Concentrating data talent in a central team has its advantages: it creates a critical mass of skills, which is good for learning and allows for more specialisation (see above). It also prevents analysts from becoming too wedded to the needs of a single part of the organisation, and encourages standardisation of tools and of data.

The major risk is that the central unit becomes isolated from the rest of the company, restricting access to knowledge and insight from other departments, and in extreme cases developing an ivory tower mentality.

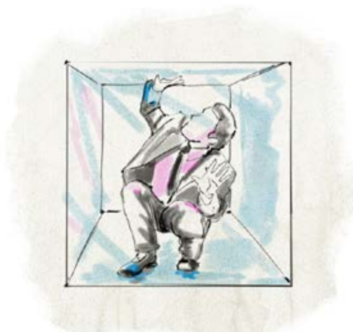
> "I don't think data is useful when it's shoved into a corner next to HR. The data people start believing the data is right. Psychology, entertainment, human gut instinct are as important to promoting success with games. The designers start seeing data as the thing their enemies use to get them to do a thing they don't want to do. My operational preference is for the data people to be embedded in the design team or the game team itself. The statisticians need to have an understanding of what their data is going to be used for."
>
> **(Creative Media Independent Consultant, CM8)**

The businesses we interviewed employed a number of strategies to avoid this, including:

- Putting business-savvy data analysts in charge of identifying and feeding business questions to the data team, and of communicating analytical results to the rest of the company.
- Rotating data talent through other departments to build their domain expertise and networks, and also seconding domain experts from other departments to the data analysis team.
- Creating tools to democratise access to data across the company, so that other functions can self-service routine data needs.[49]

If you have embedded data teams, beware of creating silos

One advantage of embedding data talent in departments across the company e.g. in Marketing or in Logistics functions, is that analysts have more direct exposure to the burning questions for that department, and are closer to the data needed to answer those questions.



But embedded data talent comes with its own challenges, not least the risk of data fragmentation and siloed thinking:

> One person would not be able to move from one department to another without there being some sort of learning curve, even though they are using what's essentially the same data; the fact it's structured differently means there's a learning curve, even though they are using the same tools.
> **(Large Financial Services Company, FS4)**

The companies we interviewed try to avoid this by setting up communities of practice for data analysts across the organisation, and defining central standards and good practices for data analysis.

> We're taking steps to create the community of practice to actually help them all the follow best practice even though they're dispersed into the business.
> **(Large Financial Services Company, FS3)**

IV. Playing and doing

A consistent theme in our interviews is that data analysts (and data scientists in particular) are often passionate about what they do: beyond financial rewards, they are motivated by the opportunity to discover new and interesting patterns in data, and to gain recognition among their peers.[50] Companies who want to retain their data talent – an important consideration in today's labour market – are looking for ways to harness these motivations, while at the same time creating business value. They do this by:

Empowering data talent

A strong finding in the literature is that individuals in creative occupations perform best when they have a feeling of ownership over their projects, and data analysts are no exception.[51] The companies that we interviewed variously give their data analysts opportunities to develop their own project ideas, run internal competitions, and 'mix things up', offering project variety and challenges to test analysts' skills and create opportunities for learning. As part of this, some companies try to replicate the workplace culture of 'startups'

> Our data scientists are motivated by freedom, a culture of moving fast and breaking things, and the opportunity to do high impact, interesting work.
> **(Small ICT Company, ICT1)**

> The main motivation is that we are that startup environment. It's a lean team. Everyone needs to pull their weight; we have very open communication. We have a good atmosphere and culture, and that's self–reinforcing.
> **(Medium–sized Financial Services Company, F7)**

Channelling talent into activities that create business value

The aforementioned freedoms need to be balanced against the imperative for data analysts to deliver business value.

> “It is important to give data scientists freedom; there is tension, though, between ‘academic’ freedom to research (doing R&D), and achieving narrow business objectives, which requires thinking of data scientists as a cog in the machine.”
> **(Large ICT Company, ICT5)**

One risk – intensified by the difficulties finding business-savvy data analysts – is that analysts get carried away by the data and go down ‘rabbit holes’:

> “Great data analysts are rarely great completer/finishers, because they just like looking in more and more detail, so you need a defined scope on a piece of work, what the objectives of it are, and what the timescales are, in order to make sure that what you get out at the end is exactly what you set out to achieve at the beginning, rather than random down a rabbit hole type route that one of the analysts might have taken.”
> **(Large Financial Services Company, FS4)**

Close communication with other departments to ensure that data analysts are addressing the questions that matter is one way to manage this risk.

Another solution is to create high levels of visibility about the impacts of data analysis on the bottom-line in order to motivate data analysts to create value. As one of our respondents put it, ‘data analysts can be highly competitive (you have to be to do a PhD), and often motivated by metrics’.

Accepting and managing failure

Data analysis projects (particularly where they use big data) are frequently innovative, and therefore risky. The companies that we interviewed were in general stoic about the need to accept and learn from project failure, but were also keen to put in place practices to insulate the businesses from the downside risks when they (inevitably) materialise. These include defining key project parameters, including goals and duration, early on, explicitly framing data projects as pilots or prototypes, and using formal development methodologies like ‘agile’ or lean production.

> “At the beginning of a project, the team will collectively examine the business problem and brainstorm what techniques to use in order to generate insights. We maintain a discipline in terms of ensuring ROI for projects, and stopping data scientists ‘going down rabbit holes’.”
> **(Large Creative Media Company, CM7)**

> “We are looking for problems with a big data solution. We identified a set of questions and turned them into pilots. There is always the risk of project failure; even if the project doesn’t fail, there is still a ‘so what’ question. We contained the risk by keeping pilots short and focused.”
> **(Large Pharmaceuticals Company, P4)**

6. Conclusions

At the beginning of this report we posed three questions. What are the data analysis skills needs of UK companies? How should companies find, manage and retain their data talent? And how do policy and education need to change to improve the supply of data talent in the UK? Our interviews have revealed wide variation in practice amongst businesses which serve to highlight the trade–offs that managers face when answering these questions, and to pinpoint where policy action is needed.

We have argued that businesses making use of data analysis can be characterised as operating in one of four distinct – albeit, in some cases, overlapping – data modes, depending on the sources of data they use and on how they apply their data. As many as one in four of the businesses represented in our interviews can be classed as working in a Big Data mode, and this proportion is growing.

The fusion of data analysis and computer programming skills that are demanded by Big Data work is giving rise to the burgeoning profession of the ‘data scientist’. In fact, all the practitioners we interviewed, regardless of data mode, were in one way or the other looking to recruit talent with this skill set.

This has allowed us to specify a profile for the ‘perfect’ data analyst in the eyes of firms. This talent has core analytical and computing skills, strong awareness of business issues, is highly creative and curious, has the ability to transform analytical insights into compelling business propositions and is a strong team–player. It should come as no surprise therefore that this talent is in short supply, with as many as four in five of businesses represented in our study reporting difficulties in recruitment.

How can the growing skills shortage be addressed?

In the near term, policymakers should work with professional bodies, Sector Skills Councils, universities and LEPs to develop CPD offers to upgrade the skills of the existing UK data workforce. Professional bodies should also do what they can to support the development of data analysis communities of practice and the ‘data analyst’ as a profession.

Policymakers also need to ensure that wider changes in immigration policy do not restrict UK employers from employing highly skilled data talent from overseas, precisely when skills shortages are becoming more severe.

Universities and industry are already making efforts to bring higher education in this area closer into line with what businesses need. Such efforts need significantly stepping up, however, and again Sector Skills Councils and bodies like the NCUB have an important role to play. UK universities are not alone in facing such challenges, and much can be learned from good practice in other countries.

In the longer term, the UK needs to strengthen its education system to improve numeracy and data handling skills across the board – including in schools – and promote the benefits of a career in data analysis to young people. It is critical to find ways of communicating the creative aspects of data analysis work, and its relevance across the economy, from drug discovery in pharmaceuticals to audience insight in creative media.

What tactics are businesses using to deal with the challenges of finding data talent, and organising it to create value?

The companies we interviewed are deploying a variety of tactics to manage these difficulties. They are actively building their reputations as ‘places to be’ for data talent, participating in online communities, making contributions to open source projects and engaging in data prediction competition platforms. In many cases, they are working more closely with universities. In all cases, they are using rigorous procedures to test candidates’ technical skills when recruiting.

However, despite all these efforts the perfect data analyst can be elusive. Businesses are therefore recruiting workers with complementary specialist skills to create teams with the required mix of skills. A diversity of skills is good for problem–solving too and creates learning opportunities for team members – which is a key motivator for data talent (though, as in other new multidisciplinary fields, firms are grappling with the challenge of creating common languages, standards and work cultures as a result).

More generally, as with other highly intrinsically motivated individuals, data scientists need to be empowered to make their own decisions based on their insights (echoing our findings from our earlier Datavores research),[52] given high levels of visibility for their work and allowed to take risks.

The interviews in this study have been with practitioners working at the coalface of the UK’s data economy. The insights will be of interest, however, to all businesses (as well as public sector bodies and not–for–profit organisations) considering how to develop their capabilities for data analysis in a labour market where talent is increasingly in short supply.

Next steps

This report was based on 45 qualitative interviews that have given us a rich understanding of the demands and realities of data analysis work in a sample of UK businesses. As a follow up, we are collecting survey and financial data from a much larger sample of companies to quantify many of the issues described in this report – including the perceived severity of skills shortages, sources of data talent, analytical methods used, and management practices adopted to create value from data talent. We will report our findings later in the year.

Endnotes

1. Manyika et al. 2011. Mayer-Schönberger and Cukier 2013.
2. Hasan Bakhshi, Albert Bravo-Biosca, and Juan Mateos-Garcia 2014.
3. By 'datavores' we mean those companies that are making a sophisticated use of data in their in their decision-making (Bakhshi and Juan Mateos-Garcia 2012) (Brynjolfsson, Hitt, and Kim 2011).
4. Just considering the period between 1900 and 1994 (when Netscape Navigator was released, arguably taking the World Wide Web into the mainstream) we find William Sealy Gosset advancing the field of statistics while working at the Guinness Brewery in Dublin; IBM researcher H.P. Luhn proposing the idea of a Business Intelligence systems, an "automatic method to provide current awareness services to scientists and engineers" (1958); John Andrew McQuown establishing Wells Fargo Management Sciences Division to model financial markets and execute trades using computers (1964); and Grant Harrison working with Clive Humby to launch Tesco's ClubCard loyalty programme, partly to track customer behaviour in order to personalise promotional offers and monitor market trends (1994). (Luhn 1958; Bernstein 2012.)
5. Loukides 2010. Patil 2011.
6. Conway 2010. Provost and Fawcett 2013. Patil 2011. Schutt and O'Neil 2013.
7. According to (Manyika et al. 2011), by 2017 the US will suffer a shortage of 'deep data talent' of up to 190,000 people. (Harlan D. Harris, Murphy, and Vaisman 2013) claim that 80 per cent of the new data scientist jobs created in the US between 2010 and 2011 have not been filled. Almost two-thirds of the UK managers surveyed by Teradata in 2013 believe there are skills shortages around big data/data science/data skills (Teradata 2013). Research by e-Skills UK shows that 57 per cent of UK companies recruiting people with 'big data skills' experience difficulties finding the right talent (e-Skills UK 2013).
8. Harris et al. 2014; Gill Press 2014.
9. Davenport 2014.
10. Cosma Shalizi 2012.
11. e-Skills UK 2013.
12. e-Skills UK 2013. Kandel et al. 2012.
13. Unsurprisingly, these are the same data features that were originally used to define 'big data' (Laney 2001).
14. Watson and Wixom 2007.
15. Booz Allen Hamilton 2013.
16. In the quotes that follow, we use an abbreviation of the sector (e.g. FS for financial services), and a number to distinguish between respondents.
17. Bell, Hey, and Szalay 2009. JISC 2014. National Research Council (US) Committee on Frontiers at the Interface of Computing and Biology 2005.
18. BBSRC 2014.
19. Manzi 2012.
20. Dhar 2012.
21. Varian 2013.
22. Harlan D. Harris, Murphy, and Vaisman 2013. Dhar 2012. Wladawsky-Berger 2014.
23. Booz Allen Hamilton 2013 contains a detailed list of the analytical techniques used.
24. Patil 2011; Loukides 2010.
25. See Patil 2011 for an overview of the technologies they use. King and Margoulas 2014.
26. Booz Allen Hamilton 2013; Dhar 2012.
27. Our interviewees were not randomly drawn from the population of UK businesses, so obviously these proportions are not representative of the wider sectors where these companies operate.
28. It is important to note that our allocation of companies into data modes was based on the experience and activities of our interviewees – there is scope for variety in data modes inside the same company. For example, pharmaceuticals companies do Data-Intensive Science in drug discovery, but might use Business Intelligence elsewhere in the business e.g. in logistics, or marketing.

29. See also (Dhar 2012) on the issue of 'problem isomorphism' across industries.
30. Patil 2011.
31. Shah 2014.
32. This number is based on those interviewees who were asked specifically about skills shortages (30). Although interviewees in the rest of our sample did in some cases mention shortages, we did not feel that their responses were strictly comparable and therefore decided to exclude them from the analysis.
33. Tambe 2013.
34. SocDM 2014.
35. http://data2discovery.org/about/
36. Ahalt 2012.
37. Rafols et al. 2012; Docherty 2014.
38. http://www2.warwick.ac.uk/fac/sci/statistics/research/csmlnetwork/
39. http://www.nuffieldfoundation.org/q-step
40. http://www.nuffieldfoundation.org/sites/default/files/files/Is%20the%20UK%20an%20Outlier_Nuffield%20Foundation_v_FINAL.pdf
41. Porkess, R. and Dudzic, S. (2013) 'A world full of data: Statistics opportunities across A–level subjects.' Available from:http://www.rss.org.uk/uploadedfiles/userfiles/files/A-world-full-of-data.pdf
42. http://schoolofdata.org/
43. Of course, it can be argued that the rising wage premium for data analytics will take care of this itself, but it can take a long time for such messages to get from the marketplace into the minds of school career advice centres, parents, and young people. Therefore, some kind of industry action to circumvent this lag effect would be helpful.
44. Amabile 1997; Sutton 2001; DeFillippi, Grabher, and Jones 2007.
45. Booz Allen Hamilton 2013.
46. Laurianne McLaughlin 2014; Patil 2011.
47. Kandel et al. 2012.
48. Booz Allen Hamilton 2013.
49. Patil 2011.
50. Patil 2011; Cathy O'Neil 2014.
51. See references in endnote 44.
52. Bakhshi, Bravo–Biosca and Mateos–Garcia op. cit.

Bibliography

Ahalt, S. C. (2012) 'Establishing a National Consortium for Data Science.'
file:///Users/juanmateos-garcia/Desktop/Read/NCDS-Consortium-Roadmap_July%20(1).pdf

Amabile, T. M. (1997) Motivating Creativity in Organizations: On Doing What You Love and Loving What You Do. 'California Management Review.' 40 (1).
http://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=00081256&AN=9712191746&h=4Bmpa0afuqTpNg%2BfCDtBLq9cAgjRp20ouEk4WhaJ1mmLSNv3oYmJ9W2oZo6qMw9iRDHeHSyU4gNYqXAVD7ocVw%3D%3D&crl=c

Bakhshi, H., Bravo-Biosca, A. and Mateos-Garcia, J. (2014) 'Inside the Datavores: Estimating The Effect Of Data And Online Analytics On Firm Performance.' London: Nesta.

Bakhshi, H. and Mateos-Garcia, J. (2012) 'Rise of the Datavores.' London: Nesta.

BBSRC (2014) 'The Fourth Paradigm of Scientific Knowledge Generation – Data-Intensive Science | Professor Douglas Kell's Blog.' Accessed 8 May.
http://blogs.bbsrc.ac.uk/index.php/2009/03/the-fourth-paradigm-of-scientific-knowledge-generation/

Bell, G., Hey, A. and Szalay, A. (2009) Beyond the Data Deluge. 'Science.' 323 (5919): 1297–98. doi:10.1126/science.1170411.

Bernstein, P. L. (2012) 'Capital Ideas: The Improbable Origins of Modern Wall Street.' Hoboken NJ:John Wiley & Sons.

Booz Allen Hamilton (2013) 'The Field Guide to Data Science.'
file:///Users/juanmateos-garcia/Desktop/Read/The-Field-Guide-to-Data-Science.pdf

Brynjolfsson, E., Hitt, L.M and Kim, H.H. (2011) 'Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance?' SSRN Scholarly Paper ID 1819486. Rochester, NY: Social Science Research Network.
http://papers.ssrn.com/abstract=1819486.

Conway, D. (2010) 'The Data Science Venn Diagram.' Drew Conway. 10 September.
http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram

Davenport, T. (2014) It's Already Time to Kill the 'Data Scientist' Title. 'CIO Journal.'30 April 2014.
http://blogs.wsj.com/cio/2014/04/30/its-already-time-to-kill-the-data-scientist-title/

DeFillippi, R., Grabher,G. and Jones,C. (2007) Introduction to Paradoxes of Creativity: Managerial and Organizational Challenges in the Cultural Economy. 'Journal of Organizational Behavior.' 28 (5): 511–21.

Dhar, V. (2012) 'Data Science and Prediction.' SSRN Scholarly Paper ID 2071041. Rochester, NY: Social Science Research Network.
http://papers.ssrn.com/abstract=2071041

Docherty, D. (2014)"Employers Must Help Universities Deliver Interdisciplinary Skills. 'The Guardian.' Accessed June 2.
http://www.theguardian.com/higher-education-network/blog/2012/may/08/developing-interdisciplinary-skills.

Dunbill, E. (2014) Defining Big Data. 'Forbes.' Accessed 7 May.
http://www.forbes.com/sites/edddumbill/2014/05/07/defining-big-data/

e–Skills UK (2013) 'Big Data Analytics Adoption and Employment Trends, 2012–2017.' London: e–Skills.

Harris, H. D., Murphy, S.P. and Vaisman, M. (2013) 'Analyzing The Analysers.' Sebastopol CA: O'Reilly.

Harris, J. G., Shetterley, N., Alter, A.E. and Schnell, K. (2014) 'It Takes Teams to Solve the Data Scientist Shortage.'
http://blogs.wsj.com/cio/2014/02/14/it-takes-teams-to-solve-the-data-scientist-shortage/

JISC (2014) 'The Data Deluge: Preparing for the Explosion in Data.' JISC Briefing Papers

Kandel, S., Paepcke, A., Hellerstein, J. and Heer, J. (2012) Enterprise Data Analysis and Visualization: An Interview Study. In 'IEEE Visual Analytics Science & Technology (VAST)'.
http://vis.stanford.edu/papers/enterprise-analysis-interviews

King, J. and Margoulas, R. (2014) '2013 Data Science Salary Survey.' Sebastopol CA: O'Reilly.

Laney, D. (2001) '3D Data Management: Controlling Data Volume, Velocity, and Variety.' Meta Group.
http://blogs.gartner.com/doug–laney/files/2012/01/ad949–3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Loukides, M. (2010)'What Is Data Science?' O'Reilly Radar.
http://radar.oreilly.com/2010/06/what-is-data-science.html.

Luhn, H. P. (1958) A Business Intelligence System. 'IBM Journal of Research and Development.' 2 (4): 314–19. doi:10.1147/rd.24.0314.

Manyika, J., Chui,M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. and Byers, A.H. (2011) 'Big Data: The next Frontier for Innovation, Competition, and Productivity.' McKinsey and Company.

Manzi, J. (2012) 'Uncontrolled: The Surprising Payoff of Trial–And–Error for Business, Politics, and Society.' New York NY: Basic Books.

Mayer–Schönberger, V. and Cukier, K. (2013) 'Big Data: A Revolution That Will Transform How We Live, Work, and Think.' Boston MA: Houghton Mifflin Harcourt.

McLaughlin, L. (2014) Big Data Hiring: 5 Facts From The Field. 'InformationWeek.' Accessed 8 May.
http://www.informationweek.com/big-data/big-data-analytics/big-data-hiring-5-facts-from-the-field/d/d–id/1113888

National Research Council (US) Committee on Frontiers at the Interface of Computing and Biology (2005) 'Catalyzing Inquiry at the Interface of Computing and Biology.' Edited by John C. Wooley and Herbert S. Lin. The National Academies Collection: Reports Funded by National Institutes of Health. Washington (DC): National Academies Press (US).
http://www.ncbi.nlm.nih.gov/books/NBK25462/

O'Neil, C. (2014) Data Science: Tools vs. Craft. 'Mathbabe.' Accessed 7 May.
http://mathbabe.org/2011/10/04/data-science-tools-vs-craft/

Patil, D. J. (2011) 'Building Data Science Teams.' O'Reilly Radar.
http://radar.oreilly.com/2011/09/building-data-science-teams.html

Press, G. (2014) Big Data Debates: Individuals Vs. Teams. 'Forbes.' 11 February.
http://www.forbes.com/sites/gilpress/2014/02/11/big-data-debates-individuals-vs-teams/

Provost, F. and Fawcett, T. (2013) 'Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking.' Sebastopol CA: O'Reilly Media, Inc.

Rafols, I., Leydesdorff,L., O'Hare, A., Nightingale, P. and Stirling, A. (2012) How Journal Rankings Can Suppress Interdisciplinary Research: A Comparison between Innovation Studies and Business & Management. 'Research Policy.' 41 (7): 1262–82.

Schutt, R. and O'Neil, C. (2013) 'Doing Data Science: Straight Talk from the Frontline.' Sebastopol CA: O'Reilly Media, Inc.

Shah, S. (2014) 'What Makes the Perfect Data Scientist?' 9 April.
http://www.computing.co.uk/ctg/analysis/2339101/what-makes-the-perfect-data-scientist

Shalizi, C. (2012) 'No, Really, Some of My Best Friends Are Data Scientists.' Three–Toed Sloth.
http://vserver1.cscs.lsa.umich.edu/~crshalizi/weblog/925.html.

SocDM (2014) 'SOCDM NEWS.'
http://www.socdm.org/images/Newsletters/SocDM_News_Issue01.pdf

Sutton, R. I. (2001) The Weird Rules of Creativity. 'Harvard Business Review.' 79 (8): 94–103.

Tambe, P. (2013) Big Data Investment, Skills, and Firm Value. 'Skills, and Firm Value.' (8 May 2013).
http://www.datascienceassn.org/sites/default/files/Big%20Data%20Investment,%20Skills,%20and%20Firm%20Value.pdf

Teradata (2013) 'Data Science in Europe – Skills Gap or Emerging Talent?'

Varian, H. (2013) 'Big Data: New Tricks for Econometrics.'

Watson, H.J. and Wixom, B.H. (2007) The Current State of Business Intelligence. 'Computer.' 40 (9): 96–99. doi:10.1109/MC.2007.331.

Wladawsky–Berger, I. (2014) 'Data Science: From Half–Baked Ideas to Data–Driven Insights.'
http://blogs.wsj.com/cio/2014/04/11/data-science-from-half-baked-ideas-to-data-driven-insights/

Nesta...

Nesta

1 Plough Place
London EC4A 1DE

research@nesta.org.uk
@nesta_uk
www.facebook.com/nesta.uk

www.nesta.org.uk

ISBN 978-1-84875-154-5

July 2014