

Dynamics of data science skills

How can all sectors benefit from data science talent?

Dynamics of data science skills: How can all sectors benefit from data science talent?

Issued: May 2019 DES5847

ISBN: 978-1-78252-395-6

The text of this work is licensed under the terms of the Creative Commons Attribution License which permits unrestricted use, provided the original author and source are credited.

The license is available at:

creativecommons.org/licenses/by/4.0

Photographs are not covered by this license.

This report can be viewed online at:

royalsociety.org/dynamics-of-data-science-skills

Contents

Foreword by Professor Andrew Blake FREng FRS	4
Executive summary	6
About this report	8
Our vision and recommended actions	9
Introduction	17
Data science demand: what the data tells us	25
Dynamics of data science: career paths and talent flows between sectors	39
Area for action: Developing foundational knowledge and skills	47
Area for action: Advancing professional skills and nurturing talent	55
Area for action: Enabling movement and sharing of talent	65
Area for action: Widening access to data in a well-governed way	75
Conclusion	85
Acknowledgements	88
Glossary	93
Data appendix	94

Foreword



Professor Andrew Blake
FEng FRS

In the 40 years since I began a doctorate in artificial intelligence vision at the University of Edinburgh, AI has changed out of all recognition. Where once a costly computer would spend hours in contemplation of one image, now a mobile phone can track faces in real time, and is backed by the immense power of the cloud to search millions of documents, and recognise speech at sometimes human-level performance. That sort of computational facility, combined with the power of statistical analysis, also gives us unprecedented analytical power to understand large and complex data sets. This combination of capabilities, that we term data science, is effecting a revolution in the way we do business, access knowledge, communicate, and understand the world.

The Royal Society has encouraged the development and use of science for the benefit of humanity since 1660. We commissioned this project, Dynamics of Data Science Skills, because we share the vision of the UK as a leading data science research nation with a sustainable flow of expertise. We believe that data science can be an exciting and fulfilling career, that also addresses society's needs. That requires the right higher education and training to be made available. More broadly, users, analysts and citizens of the future will need to be comfortable with the application of data science to societally pressing questions. This calls for data science skills to be thoroughly integrated into the school curriculum too.

Data science and engineering are growing fast and broadening in scope. No longer just the preserve of highly technical STEM- or finance-orientated roles in London, data science increasingly pervades modern business, scientific endeavour and public affairs. Children in primary school today will enter the workforce in roles that don't exist yet because of the way data, and data-enabled technologies such as artificial intelligence, are transforming the economy. This is leading to a dramatic shift in the demand for data science skills, nationwide and the need for data scientists working across all sectors. Our analysis shows that over the last five-and-a-half years there has been a sharp rise in UK job-listings for 'Data Scientists and Advanced Analysts' (+ 231%) driven predominately by increased numbers of vacancies for Data Scientists and Data Engineers.

This report provides more evidence for the case the Royal Society has already made, in *Changing education: Creating the conditions for a broad, balanced and connected curriculum*, to change post-16 education within the next ten years¹. The report also builds upon findings and recommendations of our work in data management (*Data management and use: governance in the 21st century*, with the British Academy)², artificial intelligence (*Machine Learning: the power and promise of computers that learn by example*)³, and computing education (*After the Reboot: computing education in UK schools*)⁴.

1. Royal Society 2019. Jobs are changing, so should education. See <https://royalsociety.org/~media/policy/Publications/2019/12-02-19-jobs-are-changing-so-should-education.pdf?la=en-GB> (accessed 15 April 2019).
2. Royal Society. 2017 Data management and use: Governance in the 21st century. See <https://royalsociety.org/~media/policy/projects/data-governance/data-management-governance.pdf> (accessed 15 April 2019).
3. Royal Society. 2017 Machine learning: the power and promise of computers that learn by example. See <https://royalsociety.org/~media/policy/projects/machine-learning/publications/machine-learning-report.pdf> (accessed 15 April 2019).
4. Royal Society. 2017 After the reboot: computing education in UK schools. See <https://royalsociety.org/~media/events/2018/11/computing-education-1-year-on/after-the-reboot-report.pdf> (accessed 15 April 2019).

The case to upgrade data analysis education and skills provision has also been made by other organisations, most notably in the ‘*Analytic Britain*’ briefing that was published by Nesta and Universities UK in 2015⁵.

This briefing drew on two research reports on the state of supply and demand for broad analytical skills in the UK to make recommendations spanning the whole analytical talent pipeline, including schools, colleges, universities and the labour market and industry. We hope that our report chimes with and reiterates the messages of *Analytic Britain*, whilst adding further evidence and new data to remedy skills shortages and ensure movement of talent. We also hope to continue championing the good work that is already taking place across sectors.

Our report is an extensive exploration of the current UK data science landscape. It looks at the demand for data professionals (including data analysts, data engineers, and data scientists), how this has changed in recent years, and how it varies across industrial sectors and UK regions. We use the analysis to identify four major areas for action: developing foundational knowledge and skills; advancing professional skills and nurturing talent; enabling movement and sharing of talent; and widening access to data in a well-governed way. Within these areas for action we identify priority needs and make some recommendations for addressing them.

We also share examples of exciting or innovative models and mechanisms that are already in place around the country that could be spread more widely. Those examples are available as a separate booklet, *Dynamics of data science: models and mechanisms*, that institutions and individuals can use to read about opportunities and resources in data science training and practice. A further companion booklet *Dynamics of data science: what do data professionals say about data science* collects some of the fascinating personal stories of career paths that we encountered across academia, industry, charities, and government. Our interviewees include a recent apprentice, an international entrepreneur, a physician who is also a data scientist, self-taught researchers, and data scientists working in finance, and global development. The case studies include their reflections on their careers, their experiences of moving across sectors, their observations about their role models, mentors, and professional communities, and their suggestions for improving the way things are done in data science.

Finally, I would like to say that working on this report has been a fascinating experience. The Royal Society policy staff have been a pleasure to work with, and have driven the programme with a sure touch. I also want to thank my colleagues on the steering group for their insights, and the dozens of people who have contributed at roundtables and workshops and with helpful comments and contributions.

Andrew Blake

April 2019

5. Mateos-Garcia J, Windsor G and Roseveare S. 2015 Analytic Britain: Securing the right skills for the data-driven economy. See https://media.nesta.org.uk/documents/analytic_britain.pdf (accessed 15 April 2019).

Executive summary

The skills of data scientists and engineers are in high demand. They enable organisations to extract valuable insights from data, and use them for substantial societal benefit. As data analysis methodology grows in power, and the volume of data collected increases rapidly, the number and variety of roles in data science are also growing significantly.

However, with major industry players hiring many of the most experienced data scientists and AI researchers, media reports have suggested that the natural flow of researchers from academia to industry may be reaching unsustainable levels^{6,7}. Incentives available from industry compete hard with those offered by academia, and large tech companies now even allow and encourage researchers to publish, once a particular advantage of working in universities. It is one further factor in drawing talent away from even the strongest university research groups.

There is considerable strength in UK data science in academic, industrial, charitable and government sectors. We were able to draw on the experience of representative institutions to arrive at what we hope will be helpful recommendations for employers, practitioners, and decision-makers in the private, public and university sectors. Research commissioned for this report shows an increasing need for people with data science skills, with a sharp rise in demand for Data Scientists and Data Engineers in the last five-and-a-half years. The demand spans all sectors, with specialists sought after everywhere from government departments to technology start-ups. These findings suggest that further skills gap analysis is needed to quantify the number of employed workers per opening.

Foundational training in data skills begins at school, and there is an opportunity and a need to consult broadly on a future curriculum that addresses the breadth of data skills across mathematics and science, the arts and humanities. Students need to be well informed about the ever-widening range of opportunities working with data. New hybrid forms of education in data science at school age and later, such as apprenticeships, are becoming more prevalent.

Across all sectors, those with a foundation of skills and capabilities need opportunities to deepen them and acquire new ones. Professional-level courses should be flexible and responsive. Training may need to be industry-approved and accredited, and coordination is needed between industry and universities. More informal mechanisms such as online material are also needed to allow people to (re)train through self-learning. Universities will need more than ever to retain and create the trained staff to meet these teaching needs.

-
6. Sample I. 2017 'We can't compete': why universities are losing their best AI scientists. *The Guardian*. 1 November 2017. See <https://www.theguardian.com/science/2017/nov/01/cant-compete-universities-losing-best-ai-scientists> (accessed 15 April 2019).
 7. Boland H. 2018 Britain faces an AI brain drain as tech giants raid top universities. *The Telegraph*. 2 September 2018. See <https://www.telegraph.co.uk/technology/2018/09/02/britain-faces-artificial-intelligence-brain-drain/> (accessed 15 April 2019).

Data science particularly lends itself to movement of talent between sectors, including on shorter timescales. The ability to do this will be enhanced by recognising the value of cross-sectoral working and braided careers (reciprocal arrangements that enable an individual to pursue dual, or even multiple, employment opportunities). This requires each sector to broaden its criteria and incentives to recognise and welcome more diverse forms of experience, for example academics gaining experience in the public sector or in start-ups. Attention must be paid to the needs of small and medium size businesses, charities and the public sector, addressing the challenge of large multinationals that have resources and appetite to absorb data science talent at scale. Joint positions between sectors have an important role to play in sharing talent sustainably, and in promoting diversity of experience.

Availability of data and computing power are major draws for talent. Industry dominates, offering both at considerable scale. The public and university sectors can also become more attractive to talented workers by investing more in the particular computing equipment that data science needs. Public data resources are considerable and could become uniquely compelling with more investment in curation, standardisation and careful attention to ethics and the governance of data use. In addition, access needs to be widened and opened, which also fits well with the increasing emphasis on reproducibility in research.

The challenge for UK data science is to reinforce the landscape by deploying the UK's substantial skills base flexibly, exploiting the strength of its institutions to generate and nurture more talent, and mobilising its valuable data resources ever more openly and effectively.

We are still a long way from realising the full potential of data-enabled technologies. This report highlights some of the challenges and opportunities of data science careers and some ways forward in addressing skills and mobility challenges.

About this report

This report sets out what is distinctive about data science as a discipline and offers some key statistics from our commissioned research to show the level of growth in demand for a range of data skills. We explore the different drivers and blockers around data science roles in different sectors, and highlight examples from a body of case studies of data science careers and career mobility across sectors (the complete portfolio of case studies is available as a separate publication). Engagement with data scientists, analysts and other data users informed us about some opportunities to foster talent and to enable movement between sectors. The interdisciplinary nature of data science lends itself to joint appointments, and its applied nature fits with apprenticeship mechanisms.

The following sections set out recommendations and activities across four major areas for action, with recommendations targeted across government, funders, universities, industry and the public sector to make progress towards a thriving data science landscape. They also contain a range of mechanisms for developing and sharing skills across sectors, highlighting examples of models that could potentially be replicated, scaled up and expanded.

Accompanying this report is a detailed set of case studies featuring career stories of data scientists working in a wide range of roles, levels and sectors, including Accenture, the Alan Turing Institute, Channel 4, Cambridge University, DeepZen, GCHQ, Government Digital Service, HSBC, The One Campaign, the Office for National Statistics, UCL's Institute of Neurology and the University of Warwick. (See *Dynamics of data science: what data scientists say about data science*.)

There is a clear need for collaborative, sustainable mechanisms to develop talent and this report promotes a vision for the sharing of data science talent across all sectors. We have identified a range of models and mechanisms to enable this vision, such as outreach programmes, enrichment and fellowship schemes, capability-building programmes, informal/peer-to-peer mechanisms, collaborative events and partnerships, data stores and data centres/institutes. All of which will be explored in more detail in the following chapters. The models can also be found in an accompanying document. (See *Dynamics of data science: models and mechanisms*)

Examples of good practice have been collected with the input of members of the data science community from across academia, industry and the public sector. They feature a variety of tried and tested ideas from across the UK, which require minimal to major resource support and can be led by individuals as well as institutions. The aim of the models is to inspire scale-up and cohesion.

The models and mechanisms can be used by people who are:

- concerned about the recruitment of data scientists, data analysts and domain experts;
- involved in developing data science talent at all levels;
- considering (re)training as data scientists, data analysts and domain experts;
- making decisions around skills funding on a local, regional and national scale; or
- seeking to ensure that data they hold is used for societal benefit.

Our vision and recommended actions

At the heart of this report is our vision of a healthy data science skills landscape for the UK:

VISION

The UK is a leading data science research nation with a sustainable flow of expertise. Diverse data science skills are integrated into curricula in order to develop future users, developers and citizens. Data science provides an exciting and fulfilling career choice. Data skills and appropriate infrastructure are available across sectors. Data science is applied to achieve broad societal benefit.

To achieve this vision, we focus on the four major areas for action:

- Developing foundational knowledge and skills;
- Advancing professional skills and nurturing talent;
- Enabling movement and sharing of talent; and
- Widening access to data in a well governed way.

These ideas are sketched out across the next pages and set out in more detail in the following sections, along with priority needs, models and mechanisms, and recommended actions for addressing these needs.

AREA FOR ACTION: DEVELOPING FOUNDATIONAL KNOWLEDGE AND SKILLS

NEED:

Building knowledge and skills from school level to degree level.

RECOMMENDED ACTION

Data skills for everyone

At school level, data science knowledge and skills would benefit from greater integration across the primary and secondary curriculum. Mathematics and computing communities, businesses and education professionals have a key role here.

RECOMMENDED ACTION

Support teachers to teach data skills

Develop resources, training and support for teachers and appropriate data and computing infrastructure for schools. This requires combined effort from mathematics and computing communities, businesses, and education professionals, including the new National Centre for Computing Education and National Centre for Excellence in the Teaching of Mathematics, working with the support and input of partner organisations and government departments.

RECOMMENDED ACTION

Curricula fit for the future

Post-16 curriculum change within the next ten years is vital to ensure young people leave education with the broad and balanced range of skills they will need to flourish in a changing world of work. This should start with a review into post-16 learning in the next parliament to inform future curriculum development. An analysis of the future data skills needs of students, industry and academia is needed to inform such a review.

In higher education further consideration is needed about how universities can teach data science effectively, and integrate it into university curricula as a developing and interdisciplinary set of skills and methodologies.

NEED:

Widening access to data science education.

RECOMMENDED ACTION

Raise awareness of data science careers

Data professionals work across a wide range of roles. Greater awareness of career paths could help to attract a wider pool of students. Employers could also offer work experience, host teacher Inset days and speak in schools, college and universities so that students, their teachers and careers advisers gain an understanding of possible career pathways.

RECOMMENDED ACTION

Address underrepresentation and evaluate diversity

Women make up a disproportionately small fraction of the educational pipeline associated with data science positions, and further efforts are needed by all stakeholders to address diversity, and not only of gender. This is particularly relevant as the development of data science talent needs a wider set of skills, including those involved in identifying, understanding and interpreting real-world problems. A diverse pipeline of data scientists is more likely to pick up or be concerned by inadvertent biases in algorithms that can impact on many different types of people. The Hall and Pesenti review into the growth of the UK artificial intelligence industry (2017) called for government, industry and academia to embrace the value and importance of a diverse workforce and the recommendations of this review should continue to be pursued.

AREA FOR ACTION: ADVANCING PROFESSIONAL SKILLS AND NURTURING TALENT

NEED:

Developing skills in the workforce

RECOMMENDED ACTION

Engagement between universities and employers

Universities with good industry links play a key role in developing appropriate professional training. By working in collaboration with employers they can help address regional skills gaps and productivity needs.

RECOMMENDED ACTION

Offer nimble and responsive training opportunities

Data science is fast-moving and requires innovative ways to enable the development of advanced skills. To meet the growing demand for data scientists, universities need to be agile and responsive to offer new ways of upskilling. This could potentially be achieved through MicroMasters, conversion courses and high-quality Massive Open Online Courses (MOOCs) for continued professional development.

RECOMMENDED ACTION

Develop data science as a profession

Developing a professional framework for data scientists with shared codes of practice, including appropriate governance of data collection and use and ethics training is an important short-term goal. In the longer term, professional bodies such as the British Computer Society and the Royal Statistical Society, could work with employers and universities and identify the skills needed for data scientists and consider how to address accreditation to ensure that students and professionals can be confident in the quality of new courses.

NEED

Creating the right research and working culture for data science

RECOMMENDED ACTION

Build diverse teams

Universities and the public sector in particular must work to create a culture that nurtures and retains data science talent, which can include building and supporting interdisciplinary data science teams.

AREA FOR ACTION: ENABLING MOVEMENT AND SHARING OF TALENT**NEED**

Enable movement through braided careers

RECOMMENDED ACTION

Create and fund joint positions across academia and industry

Funding bodies such as UKRI could support positions for joint appointments for a pool of the UK's most talented researchers, whose interests attract them equally to academia and industry, so that excellence can be fostered at the interface of academia, industry and government. Universities and funders should give urgent attention to enhancing mechanisms to accommodate outstanding industrial research leaders in machine learning within the academic sector.

NEED

Recognising diverse research outputs

RECOMMENDED ACTION

Commercialise research

The ways that universities encourage and support researchers in commercialising research and building spin-outs can influence researchers' abilities to hold joint appointments between industry and academia. Universities may wish to consider their strategies for research commercialisation and policies on intellectual property in order to build an environment that better supports cross-sector roles.

RECOMMENDED ACTION

Recognise diverse research outputs

Government departments and industry are likely to benefit when they enable data scientists in research roles to publish their work wherever possible; conversely, universities need to recognise the value of a breadth of experience and outputs. Alternative outputs could be recognised on academic CVs. Changes to the Research Excellence Framework that focus on institutions rather than individuals could allow universities to better recognise the contribution of data science to broader research output.

NEED

Establishing a coherent approach to policy

RECOMMENDED ACTION

Make skills a core part of the National Data Strategy

Responsibility for data policy is distributed across DCMS, GDS, Cabinet Office and DfE, but DCMS leads on delivering the National Data Strategy. This Strategy should enable departments to work closely together on data skills, building a coherent approach to delivering a healthy data science skills landscape. This will be important for the wider adoption of artificial intelligence.

AREA FOR ACTION: WIDENING ACCESS TO DATA IN A WELL-GOVERNED WAY**NEED**

Opening data and providing secure access

RECOMMENDED ACTION

Encourage data sharing where possible

Greater transparency of private sector data could help build public trust in the use of data and how it is used for decision-making purposes. The public sector could usefully consider how to widen access to its data, including sharing data, and data challenges to researchers. Journal editors should normally ensure that data is being made available to other researchers in its original form, or via appropriate summary statistics where sensitive personal information is involved. The Royal Society has published a report on Privacy Enhancing Technologies which sets out how greater use of data could potentially be enabled by PETs⁸.

RECOMMENDED ACTION

Donate data science talent

There is value in enabling data scientists to donate their time to applying data science to societal challenges. For example, through *pro bono* project work along the lines of DataKind UK, RSS Statisticians for Society and hackathons.

NEED

Providing the computing power for use by the growing data science community

RECOMMENDED ACTION

Provide access to computing power

Improving the UK's computing research infrastructure will better enable data scientists to access the necessary computing power to release the value from data and address research challenges, and will be vital for the UK to remain competitive with other countries such as the US and China. BEIS and UKRI could usefully consider the need for continuing to improve access for data scientists working across all disciplines to high-power computing, and this could helpfully be included as part of the UKRI Infrastructure Roadmap.

8. The Royal Society. 2019 Protecting privacy in practice: The current use, development and limits of Privacy Enhancing Technologies in data analysis. See <https://royalsociety.org/-/media/policy/projects/privacy-enhancing-technologies/privacy-enhancing-technologies-report.pdf> (accessed 15 April 2019).

Introduction

Data science is a rapidly developing field, and in some ways a relatively new and emerging discipline. Its development out of different disciplines, and its potential impact on society and the economy, requires a workforce with new skills. This section illustrates this critical moment and outlines our methodology for analysing how these needs can be met in practice.

Data science as a developing discipline

Making sense of data has a long history. Historically, the notion of finding useful patterns has been given a variety of names such as data mining, knowledge extraction, information harvesting and data pattern processing. It was performed by scientists, statisticians, librarians, computer scientists and others⁹. For example, in 1854, John Snow's map of cholera cases alerted him to the cause of the cholera outbreak. A cluster of dots located close to a single water supply on a map changed how we analyse and visualise data today¹⁰.

The term 'data science' emerged in the 1960s to designate a new profession that was expected to make sense of increasingly large stores of data. It wasn't until the early 2000s that the first data science journal was launched by the Committee on Data for Science and Technology (CODATA) of the International Council for Science¹¹.

Definitions of data science have evolved over time, partly as a reflection of changes in technology and data handling. In 1962, John Tukey called for a reformation of academic statistics, although there was also a move to resist it. In *The future of data analysis*, he pointed to the existence of an as-yet unrecognised science, whose subject of interest was learning from data, or 'data analysis'¹². Tukey worked between academia and industry, and provides a notable precedent for 'braided careers' in data science, working jointly at Bell Labs and Princeton University Statistics department. Over the last 50 years, statisticians, data analysts and computer scientists have played a part in the invention and development of computational environments for data analysis¹³.

A new workforce

In 2012, Harvard Business Review described the modern day data scientist as "a high ranking professional with the training and curiosity to make discoveries in the world of big data"¹⁴. D J Patil, LinkedIn Chief Scientist and author of *Data Scientist: the sexiest job of the 21st century*, explained that the focus of teams at LinkedIn was to work on data applications that would have an immediate and massive impact on the business. The term that seemed to fit best was data scientist: those who use data and science to create something new.

"A lot of what we are doing in data science involves the same computers, maths, stats, data and computationally based research that we have been doing for as long as I have been involved in research, with the current trendy label for it."

Dr James Hetherington,
Director of Research
Engineering at the Alan
Turing Institute.

9. Fayyad U *et al.* 1996 From Data Mining to Knowledge Discovery in Databases. *American Association for Artificial Intelligence* **17**, 37-54.
10. Rogers S. 2013 John Snow's data journalism: the cholera map that changed the world. *The Guardian*. 15 March 2013. See <https://www.theguardian.com/news/datablog/2013/mar/15/john-snow-cholera-map> (accessed 27 November 2018).
11. In 2013, Forbes technology contributor Gil Press put together a timeline which traces the evolution of the term 'data science' and its use, attempts to define it, and related terms. Press G. 2013 A Very Short History Of Data Science. *Forbes*. 28 May 2013. See <https://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/#30ad96c655cf> (accessed 27 November 2018).
12. Donoho D. 2017 50 Years of Data Science. *Journal of Computational and Graphical Statistics* **26:4**, 745-766.
13. Donoho D. 2017 50 Years of Data Science. *Journal of Computational and Graphical Statistics* **26:4**, 745-766.
14. Davenport T and Patil DJ. 2012 Data Scientist: The Sexiest Job of the 21st Century. *Harvard Business Review*. October 2012. See <https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century> (accessed 27 November 2018).

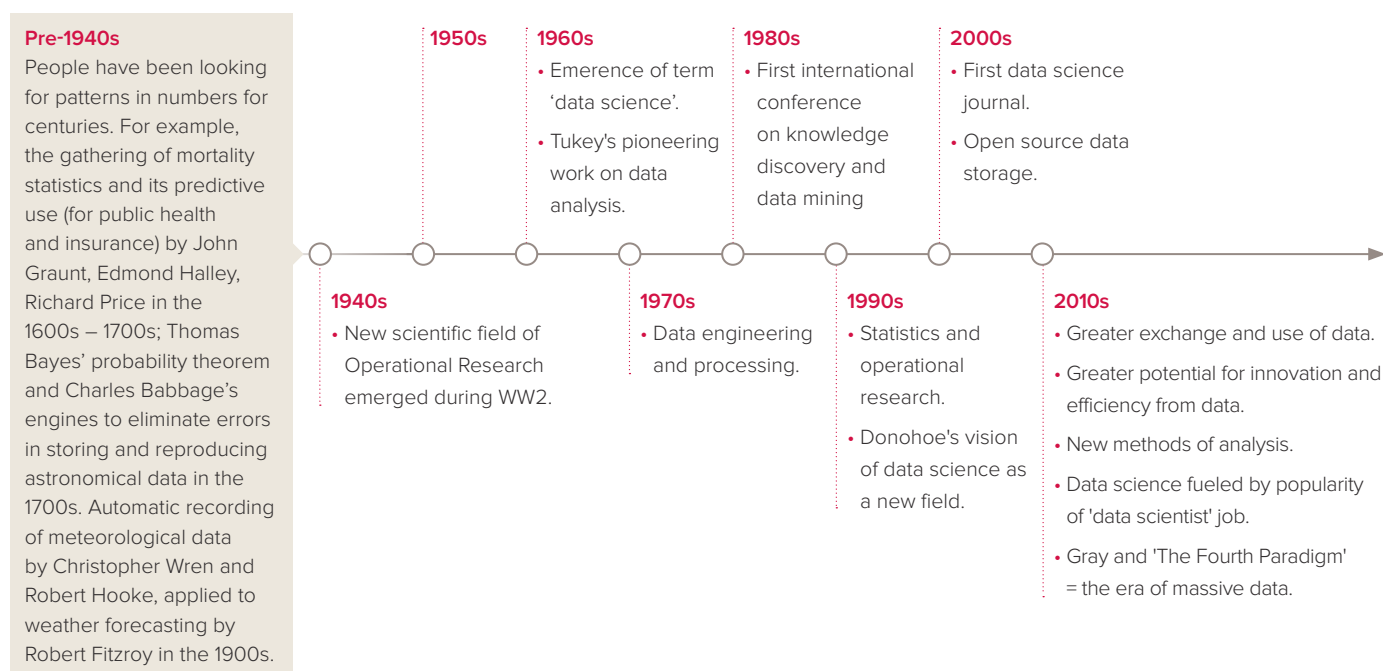
In 2007, Microsoft Professor Jim Gray coined the era of massive data the ‘Fourth Paradigm,’ stating that our capacity for collecting data has outstripped our present capacity to analyse it and so our focus should be on developing people with the skills to make sense of it¹⁵.

However, there is not yet a consistent definition of data science or the role of a data scientist¹⁶. Data science can cover a range of activities from rapid analysis of real-time data to long term evidence collection in the sciences. There is a wide variety of skills under the label ‘data science’ and people with relevant skills may associate with other disciplines.

FIGURE 1

Some key areas in data science as a discipline since the 1940s

This diagram shows some key moments and developments in the emergence of data science as an academic discipline.



15. Hey T, Tansley S and Tolle K. 2009. The Fourth Paradigm: Data-Intensive Scientific Discovery. *Microsoft Research*.

16. Press G. 2013 A Very Short History Of Data Science. *Forbes*. 28 May 2013. See <https://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/#30ad96c655cf> (accessed 27 November 2018).

Background to this study: Why now?

Greater exchange, storage and use of data:

Commercial, industrial and academic uses of data have expanded considerably over recent years. Changes to the volume, variety, and velocity of data collection have created a potentially rich resource for the digital economy. One estimate suggests that open data could help create \$3 trillion of value

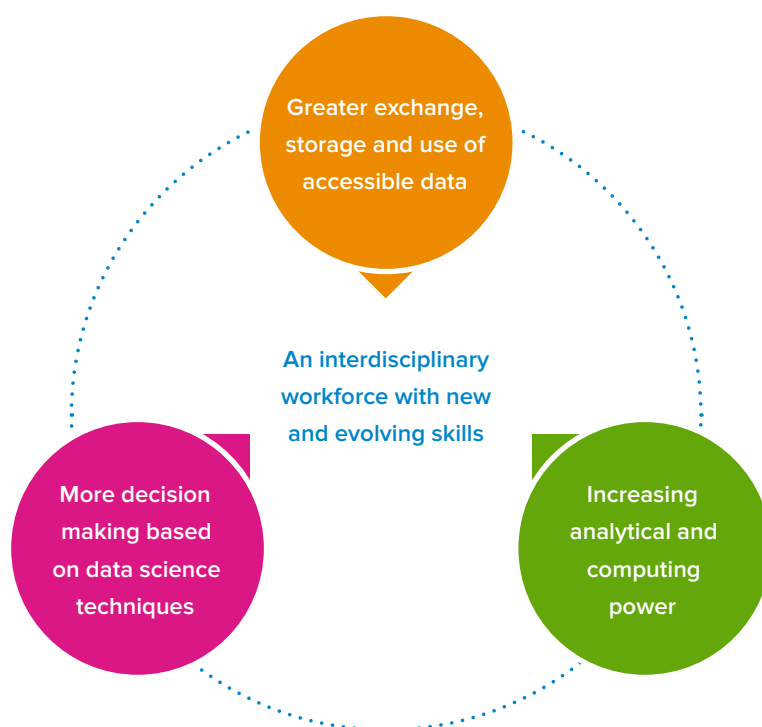
each year. In the current age, an increasing volume of information is being collected from a greater range of sources and at greater speed than ever before¹⁷. According to research from McKinsey, the volume of data continues to double every three years from digital platforms, wireless sensors and mobile phones¹⁸.

“To be a good data scientist you really do need a hell of a lot of skills”

Chanuki Illuska Seresinhe,
Senior Data Scientist,
Channel 4.

FIGURE 2

The greater exchange and use of data and the greater potential for efficiency and innovation has led to the development of a unique, interdisciplinary workforce with new and evolving skills.



17. Manyika J *et al.* 2013 Open data: unlocking innovation and performance with liquid information. McKinsey Global Institute. See <http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information> (accessed 27 November 2018).

18. Henke N *et al.* 2016 The age of analytics: Competing in a data-driven world. See <https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/the-age-of-analytics-competing-in-a-data-driven-world> (accessed 27 November 2018).

“Whether people are in the meteorological office, in the City looking at financial data, or in government statistics looking at sociological data, they have been confronted with questions and problems that go beyond traditional notions of statistics or data analysis. This has forced them to grab this notion of data science. To me, that is what really justifies its existence.”

Professor Graham Cormode, Professor in Computer Science at the University of Warwick.

Increasing analytical and computing power:

In the past, the complexity of big data limited the effectiveness of existing methods of analysis. Now, data scientists have unprecedented computing power at their disposal. In the commercial sector, analysis of data has become central to achieving competitive advantage in some sectors and has led to new applications of predictive analytics and machine learning to address business problems¹⁹. The data scientists we interviewed gave some examples of what this means for data science.

More decision making based on data science techniques:

In the public sector, combining data with analytics can, for example, allow a better understanding of population needs and help design and deliver services accordingly. Better use of data can improve the design, efficiency and outcomes of services²⁰.

The Royal Society has responded to this moment in its recent reports *Machine learning: the power and promise of computers that learn by example*, which highlighted among a range of recommendations the need for data skills at all levels – from foundational data skills to machine learning expertise at the leading edge of research²¹.

This report identified the conditions for enabling the UK to benefit from the opportunities presented by machine learning. These included the need for an amenable data environment and to generate a healthy skills pipeline from school level through intelligent users of machine learning to high level research. However, at the time of its publication there were media reports of a brain drain from academia into industry. This led the Royal Society to consider whether there are aspects of the industry/academia/public sector interface in data science that are unique compared with other sciences.

Data management and use: governance for the 21st century, produced with the British Academy, highlighted the need for new ways to govern new uses of data, including the need for responsible sharing of data and using data analysis for public benefit²². This report addresses some of the ways that these needs can be met in practice.

19. McKinsey Analytics. 2018 Analytics comes of age. See <https://www.mckinsey.com/~media/McKinsey/Business%20Functions/McKinsey%20Analytics/Our%20Insights/Analytics%20comes%20of%20age/Analytics-comes-of-age.ashx> (accessed 29 November 2018).

20. Timmis S, Heselwood L and Harwich E. 2018 Sharing the benefits: How to use data effectively in the public sector. See <https://reform.uk/research/sharing-benefits-how-use-data-effectively-public-sector> (accessed 27 November 2018).

21. Royal Society. 2017 Machine learning: the power and promise of computers that learn by example. See <https://royalsociety.org/~media/policy/projects/machine-learning/publications/machine-learning-report.pdf> (accessed 15 April 2019).

22. Royal Society. 2017 Data management and use: Governance in the 21st century. See <https://royalsociety.org/~media/policy/projects/data-governance/data-management-governance.pdf> (accessed 15 April 2019).

Workshops

To explore this further, we held scoping roundtables in 2018 to explore key questions with a range of stakeholders, including:

- what draws data science talent to particular sectors and organisations?
- how can universities and the public sector learn from industry's success?
- what kinds of models should we explore to enable a thriving landscape?
- how can we promote collaboration and cross-sector careers?

Participants highlighted a number of existing and potential new models to attract, educate and retain data science talent, and ensure a healthy research landscape.

We also explored the drivers for individual data scientists to pursue careers in different sectors and to further interrogate the models for upskilling, retaining and sharing data science talent and determine drivers for the movement of data scientists. Participants also discussed the application challenges and conditions needed for success, such as infrastructure and funding.

Statistics

The Royal Society commissioned Burning Glass Technologies, an analytics software company who provide real-time data on job growth, skills in demand, and labour market trends to provide data and analyses on UK job adverts related to data science and analytics. This was in order to determine demand trends, salary changes, location of positions, skills and experience requirements. This analysis also sets out a taxonomy of the various, interconnected occupations related to generation and use of data.

Interviews with data scientists from academia, charities, private and public sectors

To understand the drivers and blockers in more detail, a series of interviews was carried out with researchers working within the broad field of data science and particularly those who have worked across academia, government and industry. They revealed a complex network of factors that impacted on career choices. Interviewees also spoke about the mechanisms that have allowed them to thrive in multiple roles across the landscape, or move between them, and highlighted where they perceived skills gaps and gave suggestions for how to fill them.

The full set of career case studies is available as a separate publication at royalsociety.org/dynamics-of-data-science-skills. The stories illustrate the richness of the data science landscape and can be used to tell stories of data science careers.

“The pace of technology evolution is ever increasing. I think we should all be doing everything we can to enable people to be part of that. I would support training up of technical individuals who are open about intending to take careers in industry. I think that should be a strategic objective for a country.”

Dr Ilya Zheludev, Chief Data Officer at Jasmine22.

Highlights from the career stories

The full set of career case studies, *Dynamics of data science: What data professionals say about data science*, is available to download at royalsociety.org/dynamics-of-data-science-skills.

“Data science is an interdisciplinary career where you have to have a bit of maths knowledge, a little bit of coding knowledge and computers, but you also have to be quite innovative and curious about wanting to hack things.”

Alexis Fernquest, Data Scientist at the Office for National Statistics.

“I am motivated by the impact that use of data in a safe, confined, reusable, scalable way can have on the end person. I am motivated by introducing the next shift of efficiency, in terms of how you can derive a value from data. That could be anything from the ability to collect data to the ability to build a product off the back of it.”

Dr Ilya Zheludev, Chief Data Officer at Jasmine22.

“I consider myself to be a data scientist and a data science researcher, because I am still quite in love with the research side of data science. I am starting to understand that some of it might require two different skill sets, but there is definitely a lot of overlap. Data science seems to be quite multidisciplinary. You have to have a very scientific mindset and to be very logical with how you might approach your problem. But at the same time you are also interfacing with businesses and the public, so you have to be very good at communication.”

Chanuki Illuska Seresinhe, Senior Data Scientist, Channel 4.

“I wholeheartedly, absolutely, consider myself to be a data scientist. When I was a maths undergraduate, I was doing these sorts of data exercises and simple experiments. I was doing trials on friends. We would collect our opinions and guesses; run perceptual experiments and psychometric experiments, and do elementary data analyses. I loved that. But people who really like doing stuff with data did not fit into anywhere in the undergraduate curriculum at the time.”

Dr Damon Wischik, Lecturer, Department of Computer Science and Technology, University of Cambridge and former Royal Society University Research Fellow.

“My scientific background in physical sciences helped me get a leg up at the start of my career, in terms of the maths and statistical principles behind the work. A sizeable proportion of data scientists and operational researchers within government now have postgraduate degrees in physical sciences; astrophysics PhDs, particle physics PhDs. This is not an uncommon route.”

Nick Manton, Head of Data Science at the Government Digital Service and cross-government Head of Community for Data Scientists within the Digital Data and Technology profession.

“As a data scientist, I feel better when I do not know in advance how to solve something; it is more exciting. I think that is the reason that most data scientists, including me, jump from one industry to another after four or five years.”

Milton Luaces, Senior Manager at Accenture in the Fraud and Risk practice.



Chapter one

Data science demand:
what the data tells us

Data science demand: what the data tells us

The history of the data science discipline shows that it has deep roots, developing over time from a landscape of interrelated subjects. However, in recent years the demand for specialists has grown significantly, calling for new combinations of skills. There has also been an increase in the need for data skills across the board. The pervasiveness of data is rewriting the rules of many professions.

To better understand the economy's changing needs, the Royal Society commissioned the analytics software company Burning Glass Technologies to analyse the demand for data skills across the UK. Burning Glass Technologies analysed 9.2 million UK-based jobs to compare changes in demand over a five-and-a-half-year period (2013 – 2017/18). Their algorithms removed duplicates, scams and junk content from the analysis²³.

This section presents key findings from the Burning Glass Technologies data. It covers trends, salaries, skills, industry-specific demand and regional variation. Alongside this is the methodology, parameters and caveats for understanding the data, and a discussion around supply, demand and the skills gap.

See the data appendix for further information and tables.

Methodology

These findings are based on data provided to the Royal Society by Burning Glass Technologies. Burning Glass Technologies track real-time demand by collecting job postings from more than 7,500 UK online job sites to develop a comprehensive portrait. Table 1 sets out the framework categories used by Burning Glass Technologies to narrow down a set of sample occupations to include in this analysis. The table provides examples of specific occupations that fit under each of the broad framework categories and a description of the functional role.

Burning Glass Technologies have developed a taxonomy of more than 500 skill clusters by grouping skills that often travel together in job postings. In this report, clustering has been focused on the most frequently occurring skills across a range of industries, in addition to skills in emerging areas. Burning Glass Technologies use software to extract topline information about each job vacancy (such as title, employer, and industry) and 'reads' each job description to identify specific skills and qualifications that employers are seeking. A benefit of using real-time data is that the evolution of skills over time can be tracked, which is important for understanding new and developing skills such as data science²⁴.

23. Source: Burning Glass Technologies. 2019

24. Burning Glass Technologies. 2019 Frequently Asked Questions. See <https://www.burning-glass.com/about/faq/> (accessed 15 April 2019).

Parameters and caveats

There are challenges to providing definitive statistics on labour market demand and a cautious approach is recommended when interpreting the results. Some caveats to the data findings are presented below:

- **Inconsistent job titles:** Job titles are not consistent across many of these positions. An employee called a ‘Data Scientist’ at one company may have a distinctly different skill profile than a ‘Data Scientist’ at another firm, making it difficult to analyse the overall profile across all roles.
- **Missing jobs:** Many jobs are not advertised online and therefore are not included in the data, or are advertised within closed platforms such as the Civil Service portal.
- **Totals:** Not all jobs contain all information and so overall totals can differ.
- **Salary gap:** Many roles are advertised without salary details. Across all occupations in the Data Science and Advanced Analyst category, 37 – 50% of postings contained salary information in 2017/18.
- **Re-labelling of jobs:** As data science has become widely recognised some jobs are likely to have been ‘re-branded’ and this may warrant further investigation.
- **Data periods:** Two twelve-month time periods have been included in this report, covering January – December 2013 and July – June 2017/18. This represents the earliest and latest available figures at the time the data was commissioned.

Supplementing the data

Despite the caveats listed above, the data provided by Burning Glass Technologies matches reasonably well with vacancy statistics from the Office for National Statistics. Further analysis is needed to quantify the number of employed workers per opening. We suggest therefore that this data is looked at alongside alternative data sources, such as the UK labour market survey²⁵.

Another way to supplement the data provided by Burning Glass Technologies is to track emerging fields using occupations or skills as a proxy for the importance of certain roles in new industries. For example, looking at ‘Artificial Intelligence’ as a skill in order to show which roles are most commonly calling for it. This may be a good way to recognise new and emerging industries, such as those in the technology sector.

Some work has been done by other organisations to help measure the demand for emerging skills, such as AI. For example, Nesta has built a new skills taxonomy which shows the skill groups needed by workers in the UK²⁶. The taxonomy can be used as a framework to measure the demand for certain skills among employers, the current supply of those skills from workers, and the potential supply based on courses offered by education providers and employers²⁷. Their research finds that out of 143 clusters of skills, ‘Data Engineering’ stands out as the skill cluster with the highest annual median salary and growth in demand²⁸. This complements our finding that Data Engineering is in high demand.

25. Office for National Statistics. 2019 Labour market overview, UK Statistical bulletins. See <https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/bulletins/uklabourmarket/previousReleases> (accessed 15 April 2019).

26. Djumalieva J & Sleeman C. 2018 Linking skills to occupations: Using big data to build a new occupational taxonomy for the UK. See <https://www.nesta.org.uk/blog/linking-skills-to-occupations-using-big-data-to-build-a-new-occupational-taxonomy-for-the-uk/> (accessed 15 April 2019).

27. Djumalieva J and Sleeman C. 2018 Making Sense of Skills. See <http://data-viz.nesta.org.uk/skills-taxonomy/index.html> (accessed 15 April 2019).

28. *ibid*.

TABLE 1

Framework categories showing increasing levels of analytical rigour across all Data Science and Analytics (DSA) jobs.

	FRAMEWORK	FUNCTIONAL ROLE	SAMPLE OCCUPATIONS
ANALYTICAL RIGOUR	Data Scientists and Advanced Analysts	Create sophisticated analytical models used to build new datasets and derive new insights from data	Data Scientist Economist Data Engineer Biostatistician Statistician Financial Quantitative Analyst
	Data Analysts	Leverage data analysis and modelling techniques to solve problems and glean insight across functional domains	Data Analyst Business Intelligence Analyst
	Data Systems Developers	Design, build and maintain an organisation's data and analytical infrastructure	Systems Analyst Database Administrator
	Analytics Managers	Oversee analytical operations and communicate insights to executives	Chief Analytics Officer Marketing Analytics Manager
	Functional Analysts	Utilise data and analytical models to inform specific functions and business decisions	Business Analyst Financial Analyst
	Data-Driven Decision Makers	Leverage data to inform strategic and operational decisions	IT Project Manager Marketing Manager

BOX 1

Understanding the Burning Glass Technologies methodology: Classification

Burning Glass Technologies identified occupations that commonly require some mix of analytical skills and grouped them into job categories based upon similarities in skillsets and functional roles within the broader 'Data Science and Analytics' landscape. The framework is based on grouping similar occupations based on the skills and experience required. An occupation is a person's regular activity, performed in exchange for payment. Every job advert in Burning Glass Technologies' database has been assigned an occupation from the Burning Glass Technologies occupation taxonomy.

The mapping is performed by a logical rules-based system that assigns an occupation based on the information extracted from a posting. Each job advert is assigned to one occupation. The rules system compares the information found against a list of prioritised criteria, such as the clean title, the skills and certifications mentioned in the job text. It is a linearised decision tree with rules that takes the following format: 'if condition 1 and condition 2 and condition 3 [...] then outcome'. Every rule uses at least one condition which is usually based on the job title; other conditions that are based on skills, certifications, or industry are optional.

- **Framework categories:** The categories are: Data Systems Developers, Functional Analysts, Data Analysts, Data Scientists and Advanced Analysts, Analytics Managers and Data-Driven Decision Makers. Functional Analysts and Data-Driven Decision Makers are less analytical than the other four categories, and may be thought of as data-enabled, rather than pure analytics, roles. Nonetheless, they require many overlapping skillsets with other analytics roles, and are important for organisations consuming and interpreting data.
- **Occupational categories:** These six framework categories are further broken down into 40 occupations. Roles within each DSA framework category may vary in terms of required skills and experience, but are grouped together based on their function within an organisation. Data Scientists and Economists, for example, belong to the Data Scientists and Advanced Analysts category. Although these roles entail different skillsets, offer different salaries, and require different experience, employees in both are expected to create sophisticated analytical models, work with large datasets, and derive insights from data. The context in which a Data Scientist performs these tasks may differ from that of an Economist, but both roles serve similar high-level functions within an organisation and may require similar levels of analytical rigour.

Overall trends (see Table 2)

The data covers the period 2013 to 2017/18. Two time periods were compared, January – December 2013 and July 2017 – June 2018. In 2013, 6.7 million UK-based job postings were listed. In 2017 there were 9.2 million postings (an increase of 36%). Growth for all Data Science and Analytics (DSA) jobs was similar (35%).

However, during the period there is a sharp rise in job adverts calling specifically for Data Scientists and Advanced Analysts (231%). This was particularly driven by a rise in 'Data Scientists' (1,287%) and 'Data Engineers' (452%).

Salary findings (see Table 2)

The data shows that the salary changes for different categories do not necessarily correlate with growth in demand. The salary figures are indicative only, however, as many job postings do not advertise a salary, and it is possible that this is especially the case with roles at the higher end of the salary scale. See the data appendix for further information.

Skills and skills clusters

Table 3 – Skills

The most frequently occurring skills in 2017-18 were Data Science, Python and SQL. These and other skills most frequently requested for DSAA job postings were generally open source, could be applied across sectors and harness high levels of computing power.

This differed to skills most frequently requested in 2013, which included Microsoft Excel and domain-specific skills such as Economics (Appendix Table 3).

Table 4 – Skills clusters

Structuring and organising the skill-level information into clusters can help with analysis. This is particularly important in emerging areas, where jobs are less well defined.

The most frequently occurring skills clusters in DSAA job adverts were Data Science, Scripting Languages and Big Data. The Data Science skills cluster featured in 25,042 DSAA job postings (93%), suggesting high demand for this skill across occupations beyond Data Scientist.

In 2013, the most frequently requested skills clusters in DSAA job adverts were Statistics and Statistical Software skills (Appendix Table 4). Domain-specific expertise such as Economics and Medical Research were in greater relative demand in 2013 compared to 2017-18. These shifts in demand from employers likely reflect a constantly evolving data science landscape.

Regional demand (see Figures 3 and 4, and Appendix Table 2)

Regional breakdowns show the dominance of London for Data Scientist and Advanced Analysts, accounting for 58% of all postings

in 2017/18. However, for Data Scientists and Advanced Analyst job vacancies (fig 4), growth was larger (relative to the base amount) in Northern Ireland (563%), the North West (269%) and the East of England (250%).

TABLE 2

Job postings and salary information for data science and analytics jobs (including a detailed look at the Data Scientists and Advanced Analysts category).

Category	Postings			Average salary		
	2013	2017 – 18	Change %	2013	2017 – 18	Change %
Data Scientists and Advanced Analysts	8,157	27,033	231%	£52,766	£64,376	22%
Data Scientists	768	10,655	1287%	£63,053	£65,188	3%
Data Engineers	1,213	6,699	452%	£53,449	£73,559	38%
Biostatistician	936	1,823	95%	£43,601	£41,411	-5%
Financial Quantitative analyst	1,458	2,799	92%	£68,324	£71,079	4%
Economist	2,185	2,929	34%	£47,512	£49,959	5%
Statistician	1,597	2,128	33%	£44,944	£44,910	0%
Data Analysts	79,903	114,327	43%	£47,524	£46,531	-2%
Data Systems Developers	150,041	174,504	16%	£53,775	£55,126	3%
Analytics Managers	23,625	32,325	37%	£54,623	£58,835	8%
Data-Driven Decision Makers	280,218	390,328	39%	£50,264	£50,794	1%
Functional Analysts	194,667	257,745	32%	£45,606	£43,878	-4%
Total (DSA jobs)	736,611	996,262	35%			

TABLE 3

The top 10 skills listed in DSAA job adverts (2017/18).

This table shows the skills which occurred the most in 2017/18. This is measured in terms of the proportion of Data Science and Advanced Analyst (DSAA) job adverts which specified the skill as a requirement for the role. There were a total of 682 skills included in this analysis. This table displays the top ten most frequently occurring skills.

Rank	Skill	Number of DSAA job adverts requiring this skill		Percentage of DSAA job adverts requiring this skill	
		2013*	2017 – 18**	2013	2017 – 18
1	Data Science	738	11,989	9%	44%
2	Python	681	11,647	8%	43%
3	SQL	1,048	7,226	13%	27%
4	Machine Learning	352	7,089	4%	26%
5	Big Data	566	6,770	7%	25%
6	Research	2,042	5,279	25%	20%
7	Apache Hadoop	533	5,199	7%	19%
8	Communication Skills	1,843	4,849	23%	18%
9	Java	703	4,111	9%	15%
10	Scala	58	3,276	1%	12%

*8,157 job adverts included in this analysis. **27,033 job adverts included in this analysis.

TABLE 4

The top 10 skills clusters listed in DSAA job adverts (2017/18).

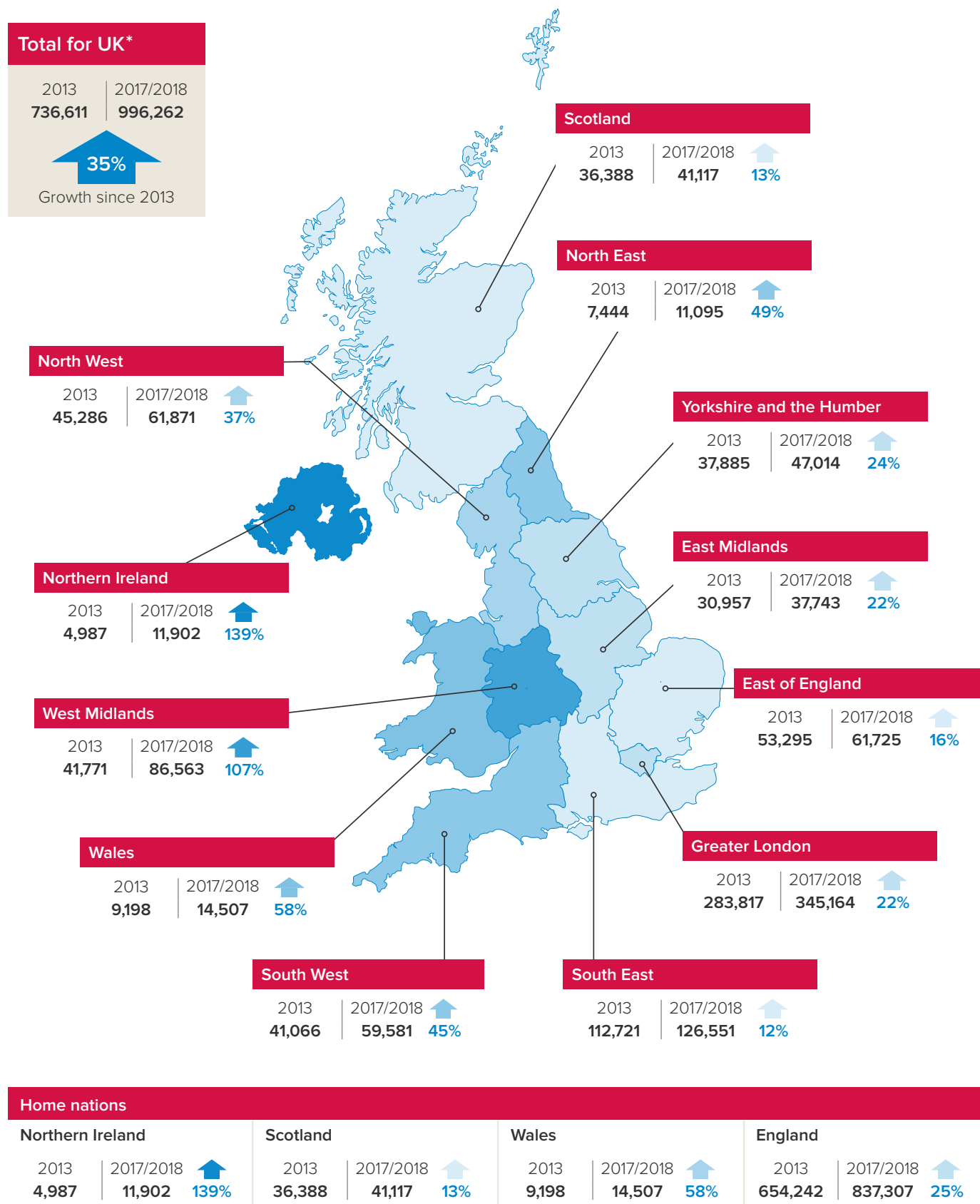
This table shows the skills clusters which occurred the most in 2017/18. This is measured in terms of the proportion of Data Science and Advanced Analyst (DSAA) job adverts which specified the skills clusters as a requirement for the role. There were a total of 278 skills clusters included in this analysis. This table displays the top ten most frequently occurring skills clusters.

Rank	Skills cluster	Number of DSAA job adverts requiring this skill cluster		Percentage of DSAA job adverts requiring this skill cluster	
		2013*	2017 – 18**	2013	2017 – 18
1	Data Science	1,056	25,042	13%	93%
2	Scripting Languages	829	23,450	10%	87%
3	Big Data	745	17,090	9%	63%
4	SQL Databases and Programming	1,151	15,532	14%	57%
5	Machine Learning	474	15,236	6%	56%
6	Data Analysis	1,628	10,636	20%	39%
7	Statistical Software	1,797	8,894	22%	33%
8	Java	703	8,258	9%	31%
9	Statistics	1,908	7,406	23%	27%
10	Software Development Principles	485	6,956	6%	26%

*Out of 8,157 job adverts included in this analysis. **Out of 27,033 job adverts included in this analysis.

FIGURE 3

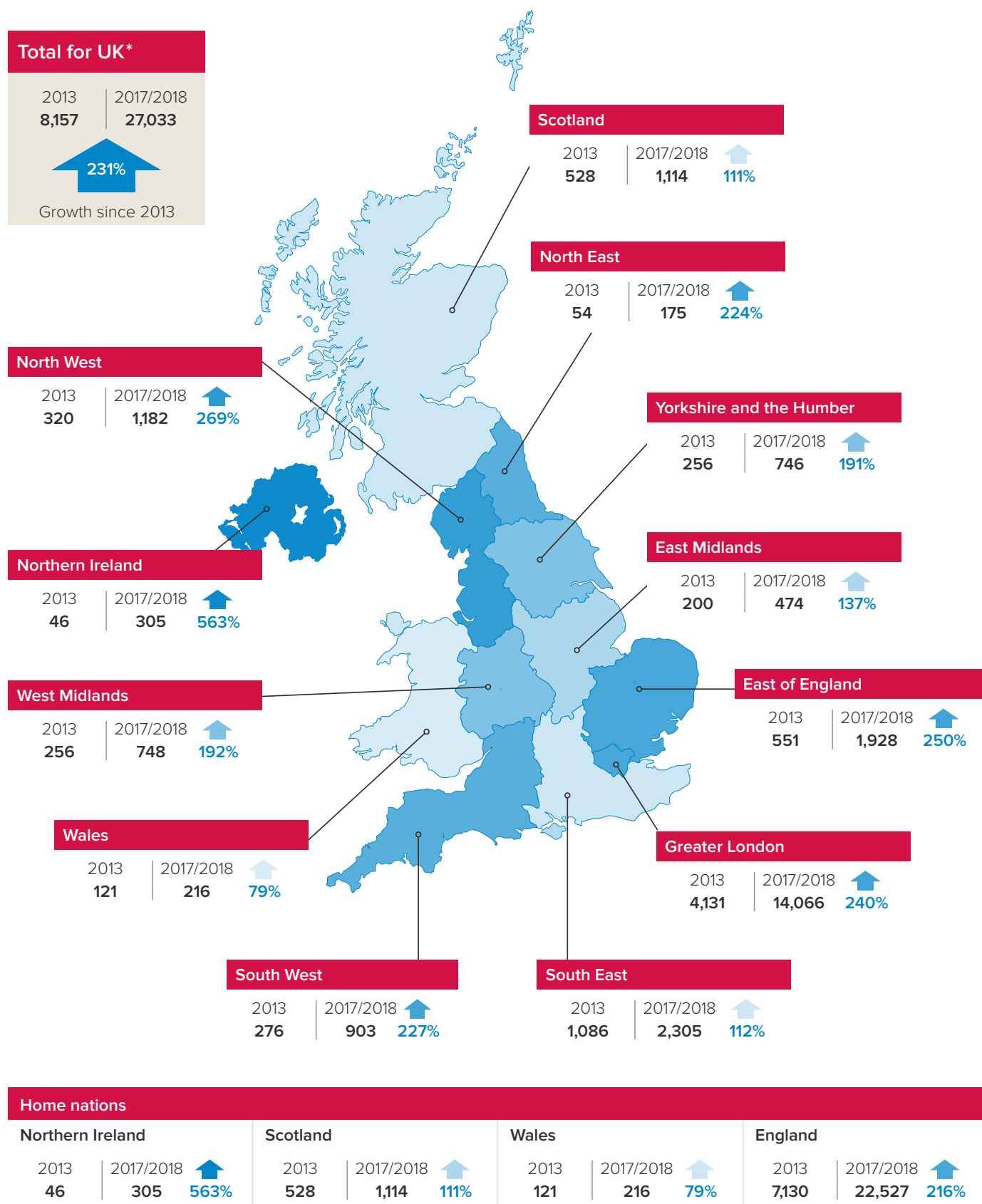
Map showing the growth of all data job postings across the UK from 2013 to 2017 – 18.



*Regional figures do not add up to the UK total because job adverts that did not include location (eg remote working) have been excluded.

FIGURE 4

Map showing the growth of Data Science and Advanced Analyst (DSAA) job postings across the UK from 2013 to 2017–18.



*Regional figures do not add up to the UK total because job adverts that did not include location (eg remote working) have been excluded.

Discussion of the data and understanding the skills gap

Since 2012, the data science industry has moved extremely quickly. It is possible that some of the jobs included in this analysis have been relabelled as ‘Data Scientist’ or associated roles, possibly leading to inflated or exaggerated growth rates. This is likely to reflect wider changes in the language used to describe data science and data scientists.

It is very hard to estimate the true gap between market demand and supply. Vacancies are not necessarily a good proxy for the number of data science jobs. One reason for this is that if job turnover increased, more vacancies will be advertised even if there has been no change in the total number of people working as data scientists. Additional analysis of the Labour Force Survey could help to aid understanding of the actual number of jobs.

There are also a number of challenges to understanding the supply of data science graduates. The Higher Education Statistics Agency (HESA) codes that are used to classify subjects studied do not include data science. In any case, only a very small minority of data scientists will come from data science courses.

Ultimately, more needs to be done to determine the existence, size and scale of the skills gap. Further analysis is needed of the way statistics on higher education are captured, so as to understand better the challenges faced by employers. The lack of traditional labour market data for these roles has created an information gap that is unhelpful to educators, employers, and policymakers as they attempt to build a workforce with the skills needed across the landscape.

The Data Skills Taskforce is currently looking at the gap between the supply of students coming through with the required knowledge and skills, and the demand required by employers. No doubt data science techniques will themselves need to be applied to predict future workforce needs.

BOX 2

The Data Skills Taskforce

The Data Skills Taskforce, chaired by Accenture and the Alan Turing Institute, sets an agenda for change to inspire, educate and upskill data talent, drawing on best practice from the UK’s leading institutions. The taskforce was established to review, promote and take forward key elements of *Analytic Britain*, across schools, universities and the labour market at large²⁹. It comprises UK businesses, data skills stakeholders and the Department for Digital, Culture, Media and Sport.

The main aims of the Data Skills Taskforce are to promote the importance of data skills, highlight critical skills gaps and monitor progress against the recommendations of *Analytic Britain*. The Data Skills Taskforce is building a platform to help SMEs develop their data science capabilities, working to quantify the UK data skills gap, which is currently significant and growing, and considering whether a data science foundation course for all undergraduates is required.

29. Mateos-Garcia J, Windsor G and Roseveare S. 2015 *Analytic Britain: Securing the right skills for the data-driven economy*. See https://media.nesta.org.uk/documents/analytic_britain.pdf (accessed 15 April 2019).



Chapter two

Dynamics of data science:
career paths and talent
flows between sectors

Dynamics of data science: career paths and talent flows between sectors

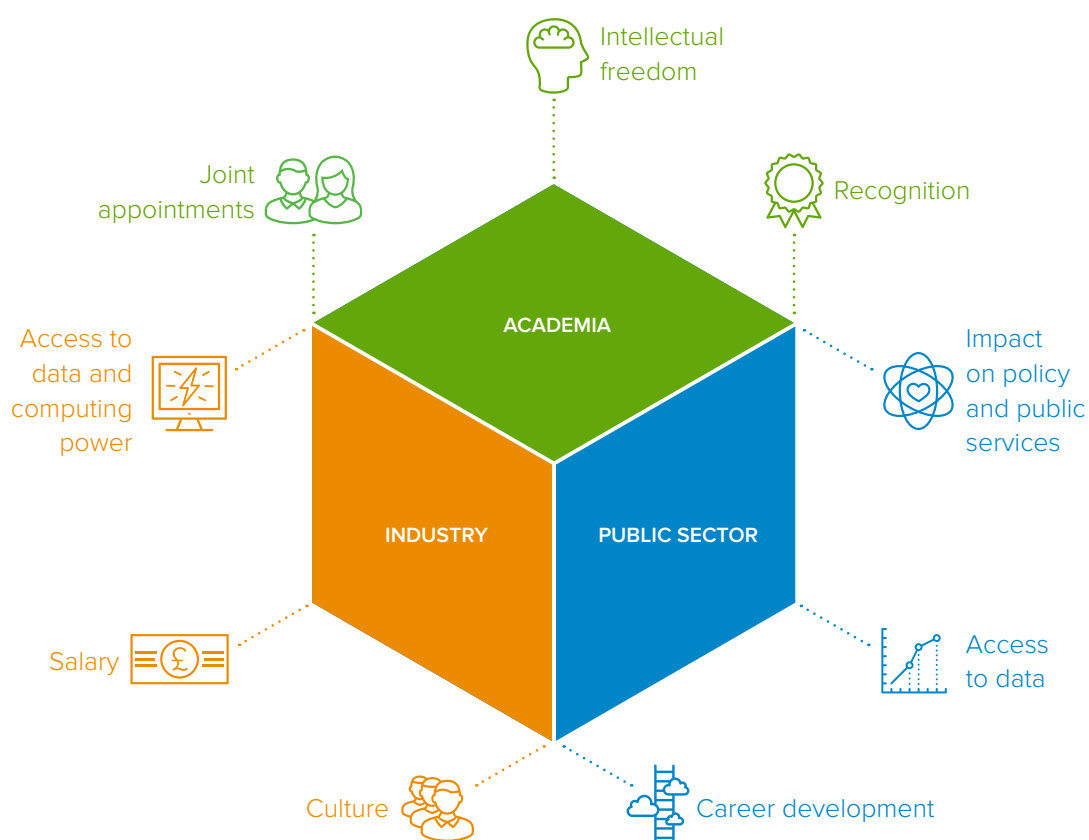
At the heart of this report is our vision for a healthy data science skills landscape. Integral to this are practical models and mechanisms, which can allow the flow of data science talent across academia, industry, government and the charity sector. This will ensure that the data science ecosystem is balanced and sustainable, with all sectors benefitting from the promise of data science capability and expertise, and an attractive and exciting diversity of career opportunities and developmental challenges for data scientists.

In this section we examine the drivers for movement of data science professionals across academia and the private and public sectors, and what can be done to facilitate greater career mobility and exchange between research in data science, and the application of data science research in the private and public sectors.

The following drivers were identified by data science professionals and those who employ them about why they might choose to move around in their careers and what might stop them from doing so. The drivers came from a series of workshops with data experts across sectors. From the list some of the key trends for each sector have been explored in more detail in a series of detailed career interviews (available in an accompanying booklet).

FIGURE 5

Examples of drivers of movement between sectors.



Note: Some of these drivers are not the exclusive pressure of any one sector but are highlighted here for illustrative purposes.

FIGURE 6

Examples of drivers and blockers by sector.

PUBLIC SECTOR: KEY DRIVERS



Societal impact

- ▶ Directly beneficial applications of data science eg improving healthcare and treatment discovery.
- ▶ Interesting opportunities to collaborate with others eg in security and defence.
- ▶ Interesting challenges that benefit society.



Access to data

- ▶ Good access to large datasets which may otherwise be inaccessible.
- ▶ Data with a high level of integrity.
- ▶ The potential to combine large datasets provides opportunity for real change.

ACADEMIA: KEY DRIVERS



Culture

- ▶ High levels of expertise and access to very talented people.
- ▶ Enabling data science through the development of theory and practice. Data scientists now 'on the bridge' and feeling valued.
- ▶ Freedom to choose research topic and research is 'open'.



Career development

- ▶ Department Heads recognising the value of data science is essential to keeping data scientists in academia.
- ▶ Spin-out culture: where universities encourage spin-outs there is an incentive to stay in academia
- ▶ Fellowships and awards create opportunities for researchers to build their career in academia.

INDUSTRY: KEY DRIVERS



Start-up culture

- ▶ Competes with academia on time to focus on science.
- ▶ Multidisciplinary teams, less siloed.
- ▶ When there is an ability to publish, industry competes with academia.
- ▶ Working directly on real-world applications (benefit to society).
- ▶ Attractive working environments at some large companies, with space for creativity.



Salary

- ▶ Salary is a pull factor towards industry plus other related financial perks, ie more permanent contracts and clearer career development paths.
- ▶ Salary is also a pull factor to international destinations, which has implications for the UK digital economy.

PUBLIC SECTOR: BLOCKERS



- Struggle to match industry/market rate salaries to bring in the right talent and skills.

ACADEMIA: BLOCKERS



- ▶ Academia provides the ability to publish, important to furthering a research career.
- ▶ Growing opportunities for joint appointments with industry.

- Struggle to match industry/market rate salaries to bring in the right talent and skills.
- Non-standard research outputs are not always valued in academia.
- Lack of clear career paths for data scientists who want to specialise.
- Access to good quality datasets.
- Lack of or limited permanent positions.

INDUSTRY: BLOCKERS



Access to data and computing power

- ▶ Data which can be used for commercial purposes.
- ▶ Large datasets can provide a competitive advantage.
- ▶ Access to infrastructure – computing power is more readily available in certain organisations.



- Research and development is subject to demand within a company and can change as strategies or economic conditions change.
- Where publishing is not an option, industry can be less attractive to data scientists.
- Salaries are not ubiquitously good across industry.

“It was feasible for me to move from industry to academia due to, in part, circumstance; the industry research environment that I was in happened to be one that allowed me to build up a strong published research portfolio. It also allowed me to do work that was public rather than a corporate secret.”

Professor Graham Cormode, Professor in Computer Science at the University of Warwick.

How to free up movement

It can be difficult for data scientists in industry to move into academia, but there are ways to make moving between sectors a natural part of the data science career path. This is important on a variety of timescales from career move to week-long study through a secondment/internship.

Publishing

In particular, there is a need to address the problem of the ‘one-way door’ out of academia which makes it difficult for researchers to return after spending time in industry or government. One enabler to this movement is the ability of researchers to keep publishing when they are outside of academia.

In 2016, Apple announced that it would begin publishing its machine learning research to help attract and retain top talent in the company³⁰.

However, publishing is not always possible outside of academia. In those cases, it is important for universities to recognise the value of work which does not result in an academic publication, and conversely, industry should recognise the skills involved in higher level research.

Outputs

Another key enabler depends on universities being more willing to recognise the value of data science experience, other than traditional published science, obtained outside universities, in the private and public sectors. Alternative outputs are increasingly recognised as an important aspect of research and should be an asset to researchers in their careers.

Experience

There is also a broader policy debate about how data science can support innovation. Researchers in fields like machine learning can have ‘a foot in each camp’ – academic jobs in involvement with start-ups or new agreements to work between sectors.

Ways of enabling people to gain experience that is valuable to all sectors can be through Fellowships and Grants that are at the interface of industry and academia. The adoption of data science is enabling government to unlock the value of the data it holds. It is being used to visualise and understand data, build tools that help policymakers access and use information, and carry out analysis that helps improve a service or drive efficiency. There are also specific challenges for movement between academia and the public sector – and fellowship models to overcome these.

30. Sheard S. 2016 Apple is finally going to start publishing its AI research. *Business Insider*. See https://www.businessinsider.com/apple-is-finally-going-to-start-publishing-its-artificial-intelligence-research-2016-12?utm_source=feedly&utm_medium=webfeeds&r=US&IR=T (accessed 15 April 2019).

Highlights from the career stories

The full set of career case studies, *Dynamics of data science: What data professionals say about data science*, is available to download at royalsociety.org/dynamics-of-data-science-skills.

“From the point of view of academic career development, my six years in Silicon Valley were completely wasted. It does not translate into anything tangible on the academic career ladder. But, I suspect that is part of why Cambridge University decided to hire me. This kind of experience must be recognised. It did not lead to papers. It did not lead to cutting-edge stuff. But it led to loads of ideas.”

Dr Damon Wischik, Lecturer, Department of Computer Science and Technology, University of Cambridge.

“Once you have left academia and you are not publishing high-quality papers, it is very difficult to go back. If there was more funding for projects that required one commercial partner and one academic partner, that would be interesting.”

Chanuki Illushka Seresinhe, Lead Data Scientist at Popsa.

“The counterpoint is that mobility can be harder in either direction if the gap is wider. It is often said that it is difficult for people from industry to move into academia if they have not maintained a research profile.

Equally, I think if you have been working in an academic environment focused on publishing and you want to go to a more industrial employer, one of the things you will face is that they will say, ‘Before we get started, can you do some programming exercises or fairly hands-on stuff?’ For someone who has been working on equations and theory, it can actually be a bit of a wrench to reorient to those more hands-on expectations that organisations have.”

Professor Graham Cormode, Professor in Computer Science at the University of Warwick.

“This is a major data science/AI challenge for government, innovative companies and universities. In my experience, the boundary between universities and government is potentially difficult: the timescales and measures of success are very different on either side. But there are various experiments, for example the Policy Fellowships pioneered by the Centre for Science and Policy, which show how the two sides can learn from each other to the enormous benefit of our society as a whole.”

Professor Frank Kelly CBE FRS, Professor of the Mathematics of Systems at the University of Cambridge.



Chapter three

AREA FOR ACTION:

Developing foundational
knowledge and skills

AREA FOR ACTION:

Developing foundational knowledge and skills

“It is not common for my generation to be able to program. There are loads of really great online resources, but I think it can still be difficult for people to see why it would be useful.”

Dr Amy Nelson, Senior Research Associate, UCL Institute of Neurology.

Education should provide a grounding to ensure that all young people develop underpinning data science knowledge and skills. The data experts that we spoke to highlighted a range of core skills and disciplines that need to be developed early on including coding, computer science, mathematics, machine learning, statistics, and more.

Last year the Royal Society published a review of how data science skills are nurtured in England's curriculum³¹. The review was written by a curriculum expert, Dr Vanessa Pittard, and it identified some barriers to embedding data skills into the curriculum as well as some opportunities for further development.

The review found that at primary level, pupils are likely to gain a reasonable introduction to underpinning elements of data science if they are taught well. However, at secondary level, the lack of systematic progression in relevant aspects of computing is a major curriculum challenge.

Last year the Royal Society also conducted a study identifying a range of initiatives that seek to increase the proportion of young people studying computing (particularly girls)³². The study, co-funded by Microsoft and Google, found that 54% of English schools do not offer Computer Science GCSE, and put recommendations in place to encourage greater take up and increase teachers' confidence. The report also found that barriers exist around teacher confidence in applying mathematics to domain questions, and in conceptual and technical aspects of computing.

Alongside the technical skills, interviewees in this project also highlighted a range of other core skills for working well with data including adaptability, curiosity, empathy, problem-solving and story-telling which supports the case for embedding data science in an array of subjects.

There are pools of potential talent which could be reached to address local needs and there could be more courses, apprenticeships and work placements outside of London and the South East. Employer-led Trailblazer groups and public sector bodies should further support and expand existing programmes, such as those run by the Office for National Statistics and the BBC, and work together to resolve knotty delivery issues.

This chapter sets out two priority needs and offers specific recommendations for developing foundational knowledge and skills for data science. It also highlights successful or innovative models and mechanisms for developing these foundations and addressing these needs.

31. Pittard V. 2017 The integration of data science in the primary and secondary curriculum. See <https://royalsociety.org/~media/policy/Publications/2018/18-07-18-The%20Integration%20of%20Data%20Science%20in%20the%20primary%20and%20secondary%20curriculum.pdf?la=en-GB> (accessed 15 April 2019).

32. Royal Society. 2017 After the reboot: computing education in UK schools. See <https://royalsociety.org/~media/events/2018/11/computing-education-1-year-on/after-the-reboot-report.pdf> (accessed 15 April 2019).

NEEDS AND RECOMMENDED ACTIONS

NEED:

Widening access to data science education

Our work on the educational data science pipeline, discussions with data scientists and analysis of job market data, showed that there is a need to follow up a number of priority needs in order to develop the foundational level knowledge and skills required for professional roles working with data. These needs cover building core skills at school level, supporting teachers, expanding apprenticeships and promoting outreach initiatives to introduce young people to data skills.

RECOMMENDED ACTION

Data skills for everyone

At school level, data science knowledge and skills would benefit from greater integration across the primary and secondary curriculum to ensure that everyone leaves school with foundational data skills. Mathematics and computing communities, businesses and education professionals have a key role in ensuring that relevant insights are built into the education curriculum and associated enrichment activity in schools, ensuring that data science skills are appropriately embedded across subjects and phases.

RECOMMENDED ACTION

Support teachers to teach data skills

Developing the resources, training and support for teachers requires combined effort from mathematics and computing communities, businesses, and education professionals, including the new National Centre for Computing Education and National Centre for Excellence in the Teaching of Mathematics, working with the support and input of government departments and other partners. This can build on important work already carried out by the Department for Education and others in this area, and examples such as the Royal Geographical Society's Data Skills in Geography programme (see Models and mechanisms).

Issues around teacher confidence could be addressed through professional development opportunities and more provision of detailed guidance. Subject specific knowledge could be delivered in collaboration with universities and professional bodies (eg subject associations). There are possibilities for incorporating real uses of data across the sciences and humanities. One area that is relatively underplayed in the current curriculum is developing investigation using GDS/GIS data which could be incorporated into humanities subjects. Appropriate computing and data infrastructure for schools will be crucial to support the teaching of data skills.

“In my apprenticeship there was a whole unit dedicated to how you work with other people in an office. You had to learn to express your opinion and convince people of your ideas. Apprenticeships could be run across different sectors so that alumni can build professional networks.”

Alexis Fernquest, Data Scientist, Office for National Statistics and former apprentice.

“I would like to see more people with developed skills in computational statistics and machine learning coming out of undergraduate degrees that are not computer science and physics. For example, the social sciences, life sciences and environmental sciences.”

Dr James Hetherington,
Director of Research
Engineering.

RECOMMENDED ACTION

Curricula fit for the future

Looking further ahead, an analysis of the future data science needs of students, industry and academia should be undertaken to inform future curriculum development. Curriculum content groups should consider the place of data science within the curricula they are developing now, ahead of the next curriculum review. In addition to the relevant areas of mathematics, computer science and data literacy, the ethical and social implications of machine learning should be included within teaching activities in related fields, such as Personal, Social and Health Education.

Post-16 curriculum change within the next ten years will be vital to ensure young people leave education with the broad and balanced range of skills they will need to flourish in a changing world of work³³. A review into post-16 learning should consider the many ways in which the post-16 curriculum could be improved and the factors which affect the options open to young people.

At university level, further consideration is needed about how universities can teach data science effectively, as a developing and interdisciplinary discipline.

NEED:

Widening access to data science education

There is also a need to address the underrepresentation of women in the talent pipeline. Across the mathematical sciences, 37.1% of university students were female in 2018. Computer science was the subject with the widest gender gap across all degree levels in 2018 (82.8% male). This was most pronounced at the undergraduate levels: just 15.1% of first degree undergraduates were female. Forbes data from a US study of technical education provider, General Assembly, found that in data science programmes/boot camps in the US female participation lags with 35.3% female students enrolled in a five month period (2016/17)³⁴.

In the UK, just 4% of the UK tech industry is from a black, Asian and minority ethnic (BAME) background, compared with 11% in the working population as a whole, at the last census. With low diversity in academic programmes and industry bootcamps, it is unlikely that diversity gaps will close in the near future without significant investment³⁵.

33. Royal Society 2019. Jobs are changing, so should education. See <https://royalsociety.org/~media/policy/Publications/2019/12-02-19-jobs-are-changing-so-should-education.pdf?la=en-GB> (accessed 15 April 2019).

34. Pricenomics. 2017 The Data Science Diversity Gap. *Forbes*. 28 September 2017. See <https://www.forbes.com/sites/priceonomics/2017/09/28/the-data-science-diversity-gap/#56ccb88c5f58> (accessed 15 April 2019)

35. Douglas R. 2018 Lack of diversity increases risk of tech product failures. *Financial Times*. 14 November 2018. See <https://www.ft.com/content/0ef656a8-cd8a-11e8-8d0b-a6539b949662> (accessed 16 April 2019).

RECOMMENDED ACTION

Raise awareness of data science careers

Data professionals work across a wide range of roles. Greater awareness could help to attract a wider pool of students. Employers could also offer work experience, host teacher inset days and speak in schools, college and universities so that students and their teachers and careers advisers gain an understanding of emerging career pathways.

We have published a range of case studies alongside this report to highlight the wide range of roles in data science and related fields.

The Institute for Apprenticeship's occupational maps are an example of how to explain career pathways to students, parents and teachers, and organisations recruiting data scientists can share stories of what they do and the impact of their work.

RECOMMENDED ACTION

Address underrepresentation and evaluate diversity

Women make up a disproportionately small fraction of the educational pipeline associated with data science positions, and further efforts are needed by all stakeholders to address the shortage. Diversity, and not just of gender, is relevant as data science talent requires a wide set of skills, including those involved in identifying, understanding and interpreting real-world problems. A diverse pipeline of data scientists is more likely to pick up or be concerned by inadvertent biases in algorithms that can impact on many different types of people. The Hall and Pesenti review into the growth of the UK artificial intelligence industry (2017) called for government, industry and academia to embrace the value and importance of a diverse workforce and the recommendations of this review should continue to be pursued.

A significant barrier to improving diversity is the lack of access to data on diversity statistics in industry and in academia. We encourage institutions to be more transparent about diversity statistics in an effort to improve diversity in the field. Funding for diversity initiatives could be expanded to augment the presence of minority backgrounds and ethnicities.

“There should definitely be more diversity in role models. It shows you that ‘this is a career for people like me.’”

Aimee, Government Communications Headquarters (GCHQ).

MODELS AND MECHANISMS

“You have to be quite resilient and open to change. In 10 years’ time it will be different again because they have only really just started trying to get coding into schools. You are going to have all these people coming out who are already excellent at coding.”

Alexis Fernquest, Data Scientist at the Office for National Statistics.

Through our workshops and interviews we identified some key models and mechanisms that can help to address these priority needs. The mechanisms include integration of data skills into the school curriculum, apprenticeships to develop practical skills, alternative degree pathways and outreach programmes. The specific examples below serve as models of what can and has worked in specific contexts.

Mechanism

Integration of data skills into the school curriculum.

Model

Data Skills in Geography programme, Royal Geographical Society with the Institute of British Geographers

Recent changes in curricula in schools and at university, along with a recognised skills gap, have brought renewed emphasis on students being trained in data skills (the collection, analysis and presentation of data) in geography at GCSE, A Level and in undergraduate courses – and within Higher Education geography has been recognised by HEFCE as a ‘part-STEM’ subject. The shift is presenting new challenges for many school teachers, particularly those with little prior experience of such skills. In response to these changes and challenges, the Royal Geographical Society (with IBG) is leading a programme Data Skills in Geography, supported by funding from the Nuffield Foundation. RGS has also established new networks and strengthened existing ones, including creating a Data Champions scheme made up of teachers who are data enthusiasts with different levels of experience.

Mechanism

Apprenticeships to develop practical skills.

Model

The Office for National Statistics: Data Science Campus

The Office for National Statistics (ONS) is leading the way in developing a degree level apprenticeship in data science on behalf of the public and private sectors. The course is one of a number of new offerings at the new Data Science Campus³⁶. The campus is based at the ONS headquarters in Newport and part of a £17 million investment in statistics and data by the UK government to modernise and improve the statistics it produces. The funding was allocated after the Bean Review highlighted concerns around the use of administrative data³⁷.

All the training programmes at the campus cover three levels: raising awareness, embedding core skills and developing expertise. The ONS set up the scheme because it felt that there were no suitable existing programmes available in the university sector. The need to develop its own courses was important to equip data scientists with better insights into the UK economy for policymakers. All the applicants to the campus come from very different and diverse backgrounds. Some are mathematicians, physicists, social scientists, and artists. School-leavers and career-changers are also welcome. The ONS wanted a diverse array of educational pathways so that everyone can fulfil their potential, realising that work-based learning has a key role in delivering that ambition.

36. Poluck M. 2017 ONS opens £10 million hub to exploit digital data. *UK authority*. 27 March 2017. See <http://www.ukauthority.com/data4good/entry/7009/ons-opens-%C2%A310-million-hub-to-exploit-digital-data#> (accessed 18 April 2018).

37. Chu B. 2015 Bean review tells ONS to be more “intellectually curious”. *Independent*. 3 December 2015. <https://www.independent.co.uk/news/business/news/bean-review-tells-ons-to-be-more-intellectually-curious-a6758226.html> (accessed 18 April 2018).

Mechanism**Alternative degree pathways.****Model****Q-Step, Nuffield Foundation, ESRC and HEFCE (now Office for Students (OfS))**

The Q-Step programme, launched 2013/14 aims to address the quantitative skills gap in social sciences in the UK. With an initial six-year £19.5 million investment from the Nuffield Foundation, the ESRC and HEFCE, around 60 new academic staff were employed to use innovative approaches to embed quantitative methods and data analyses into social science teaching. There are currently 18 universities across the UK funded to deliver specialist undergraduate programmes, including new courses, work placements and pathways to postgraduate study.

Over 1,000 undergraduates a year start a Q-Step degree pathway, whilst almost ten-times as many experience enhanced quantitative teaching by taking one or more Q-Step modules. Over 420 employers are involved with offering placements to more than 860 students, ranging from private/public to the third sector and government departments. The programme is having a notable and positive impact on the data skills of social scientists taking up postgraduate research studentships. Q-Step will continue to be funded by the Nuffield Foundation and the ESRC for two more years (2019/20 and 2020/21) and is undergoing an independent impact evaluation.

Mechanism**Outreach programmes.****Model****Code First: Girls**

Code First: Girls supports young adult and working age women to develop further personal and professional skills. Code First: Girls runs free coding courses; it also connects women to a community of other talented and like-minded women and companies who can support and accompany them through their professional development. Code First: Girls helps companies train their people, recruit new people, and develop their talent management policies and processes so they don't miss out on female tech talent.

“Q-Step academics have enriched social science teaching by integrating data skills into their courses. Most students are now handling data in R. It demonstrates that what we refer to as STEM skills don't just come from taking STEM degrees.”

Dr Simon Gallacher, Head of Student Programmes at the Nuffield Foundation.



Chapter four

AREA FOR ACTION:

Advancing professional
skills and nurturing talent

AREA FOR ACTION:

Advancing professional skills and nurturing talent

“There are four main technical skills that a data scientist needs: statistics, machine learning, data and computer science. What is in short supply is having all four of them together, plus the soft skills.”

Milton Luaces, Senior Manager at Accenture – Applied Intelligence.

Data science is fast-moving and needs innovative mechanisms to develop advanced skills quickly. When people are in employment, there are a number of needs to be met in enabling them to keep their data science skills up to date and to ensure that they can work most effectively. These include:

- developing teams and a workplace culture that enables data scientists to make the best contribution in their sector;
- enabling individuals to keep their skills up to date whether they are data science specialists or people in a range of professions who increasingly need to work with data;
- supporting career changes to meet individual, organisation and regional needs.

In the longer term, problem solving, resilience, and continuous learning will also be necessary to enable people to adapt to change, particularly as technology changes jobs and the opportunities to collect more and more data continue to grow.

There is a need to develop researchers and employees with good data engineering skills who have knowledge and experience of handling ‘big data,’ huge unstructured datasets, data sources and/or real-time data. Across all sectors there are massive, high volume and high velocity datasets that need to be stored, processed and analysed in real time, and this requires the creation and maintenance of infrastructure for big data. Typically this is work that would be undertaken by data engineers, but currently it is often done by data scientists because there are not enough data engineers to take on these roles.

New technologies are changing the roles that data scientists are performing and this is having an impact on the depth and breadth of skills that are being sought after. New technologies are performing roles that previously required highly skilled employees, making people with more junior data qualifications increasingly effective.

However, the commercial sector could usefully do more to help train and develop employees with the right skillsets. Employers have a role in upskilling the workforce by training existing employees, particularly those at risk of losing their jobs through automation, and can work with universities to co-produce training. Higher education institutions with good industry links play a key role in developing appropriate professional training. By working in collaboration with employers they can potentially address regional skills gaps and productivity needs.

“Industry can play a slightly bigger role in being involved in the training of people, regardless of what age or stage they are at. At the moment, it is an employer and employee relationship. But I can see it being a lot more collaborative and I have come across institutions that are starting to recognise this. They are moving towards a mindset of upskilling people, exposing them to their company, giving them a grounding and capabilities to help solve the problems that particular institution faces. I was sponsored by a company during my PhD and I would encourage more institutions to do that.”

Ilya Zheludev, Chief Data Officer for Jasmine 22.

In the public sector, collaboration is already helping to upskill the workforce. At the Office for National Statistics there are Joint PhD/MSc projects, or placements, for example the joint Turing-Campus PhD programme. The ONS is also involved with several of the Centres for Doctoral Training in big data, and runs CPD courses for analysts – typically short intensive courses linked to specific needs.

Beyond this, specialist skills are often developed through self-learning programmes. An increasing proportion of job adverts are calling for software skills such as Hadoop, Python and R. Informal learning and the open source movement are enabling people to develop these skills in non-traditional ways.

In 2017, Bernard Marr, writing for Forbes, put together a list of free online data science courses with links and endorsements from well-established institutions³⁸. For example:

- Data Science Specialization (Johns Hopkins University and Coursera): One of the longest-established online data science courses; and
- Data Science Essentials (EdX and Microsoft): Part of the Microsoft Professional Program Certificate in Data Science;
- Become a Data Scientist (Dataquest with endorsements from Uber, Amazon and Spotify): Independent online training provider offering three pathways (analyst, scientist and engineer); and
- Data Mining Course (KDNuggets): Well-known business and data science website that has compiled its own free data-mining syllabus.

Whilst Massive Open Online Courses (MOOCs) suffer from high dropout rates and irregular provision, they also enable people to undertake independent study and upskilling alongside other more formal training programmes.

Other informal mechanisms include peer learning such as forums and meet-ups, competitions and clusters. One example is Kaggle, an education platform that also holds competitions³⁹. The aim of meet-ups is to promote free, open, dissemination of data science knowledge. They encourage data science peer-to-peer learning and sharing, collaboration among data scientists and data start-ups. They also promote open source data science tools⁴⁰.

Some meet-ups focus on underrepresented groups, such as the London-based 'Inspiring Women in Data Science', which has almost 1,000 members and runs several events throughout the year⁴¹. In addition, there are other groups that focus on specialisms within data science including Women in ML, Black in AI, Queer in AI, LatinX in AI, the AI Club for Gender Minorities, RLadies and PyLadies, an international mentorship group for women who code in Python with chapters in London, Edinburgh and Dublin.

Community knowledge-sharing events such as panel/roundtable events on specific hot topics also bring together data scientists to discuss, share and take advantage of community knowledge.

“I started off with Tableau, storytelling and data visualisation. On the side, I was starting to learn R and Python for more of the deep dive work on the data.”

Kevin Koene, former Junior Data Scientist, The One campaign.

38. Marr B. 2017 The 9 Best Free Online Big Data And Data Science Courses. *Forbes*. 6 June 2017. See <https://www.forbes.com/sites/bernardmarr/2017/06/06/the-9-best-free-online-big-data-and-data-science-courses/#466bc8543cdb> (accessed 15 April 2019)

39. Kaggle (no date) Kaggle is the place to do data science projects. See <https://www.kaggle.com/> (accessed 15 April 2019).

40. Meetup (no date) Inspiring Women in Data Science. See <https://www.meetup.com/IWDSuk/> (accessed 15 April 2019).

41. Sekinah T. 2017 Info on Women in Data groups. *Data/Q*. 6 December 2019. See <https://www.dataiq.co.uk/blog/info-women-data-groups> (accessed 15 April 2019).

NEEDS AND RECOMMENDED ACTIONS

“If you want a senior data scientist with more than 10 years’ experience, we are short on that and we will still be short for another three or four years.”

Milton Luaces, Senior Manager at Accenture – Applied Intelligence.

Need: Developing skills in the workforce.

Through our workshops and interviews we explored the need to ensure that data science skills can be nurtured so that organisations can retain talented individuals and get the most out of their skills, and to enable them to keep those skills up to date in a fast moving field.

RECOMMENDED ACTION

Engagement between universities and employers

Universities with good industry links play a key role in developing appropriate professional training. By working in collaboration with employers they can help address regional skills gaps and address productivity needs.

Employers have a role in upskilling the workforce by training existing employees, particularly those at risk of losing their jobs through automation, and can work with universities to co-produce training. By working in collaboration with employers, universities can potentially address regional skills gaps and address productivity needs. This could involve working across professional disciplines to understand the type and level of data science skills that will be needed by professionals in fields such as law, healthcare, and finance. The Royal Society’s *Machine learning* report highlighted the need for training informed users of machine learning techniques, for example⁴².

RECOMMENDED ACTION

Offer nimble and responsive training opportunities

Data science is fast-moving and requires innovative ways to enable the development of advanced skills. To meet the growing demand for data scientists, universities need to be agile and responsive to offer new ways of upskilling.

This could potentially be achieved through MicroMasters, conversion courses and high-quality Massive Open Online Courses (MOOCs) for continued professional development.

In the short term, a strong and effective pipeline of practitioners is likely to be established through government support for advanced courses – namely Masters degrees – which those working across a range of sectors could use to develop data science skills at a high level. Encouragingly, the Office for AI, the British Computing Society and the Institute for Coding are already looking at improving the supply of skills through a new series of Masters courses, in consultation with The Alan Turing Institute. Options such as MOOCs should be considered as a vehicle for developing skills ranging from informed users through to expert data engineers.

42. Royal Society. 2017 Machine learning: the power and promise of computers that learn by example. See <https://royalsociety.org/~media/policy/projects/machine-learning/publications/machine-learning-report.pdf> (accessed 15 April 2019).

RECOMMENDED ACTION

Develop data science as a profession

Developing a professional framework for data scientists with shared codes of practice, including appropriate governance of data collection and use and ethics training is an important short-term goal. In the longer term, professional bodies such as the British Computer Society and the Royal Statistical Society, could work with employers and universities and identify the skills needed for data scientists and consider how to address accreditation to ensure that students and professionals can be confident in the quality of new courses.

Need: Creating the right research and working culture for data science.

Our interviews with data scientists also highlighted key aspects of workplace culture that can be addressed to ensure that we have the right skills where they are needed. These included the view that data scientists flourish when working together in teams. The following are recommended actions for ensuring that data scientists and others already in work can do the best work with data that they can.

RECOMMENDED ACTION

Build diverse teams

Universities and the public sector in particular must work to create a culture that nurtures and retains data science talent, which can include building and supporting interdisciplinary data science teams.

Universities in particular could recognise data science roles, create appropriate job titles and support permanent roles and career progression for data scientists to create a collaborative, cross-disciplinary culture. There are many examples of good practice, such as Research Software Engineering Groups, which provide a home for research programmers who collaborate with researchers on multiple research projects⁴³.

43. Research Software Engineers Association. 2019 Welcome to the UK Research Software Engineer Association See <https://rse.ac.uk/> (accessed 15 April 2019).

MODELS AND MECHANISMS

“At the Office for National Statistics, they have a training programme and they were willing to develop me for the next role. They also encourage us to move around every two years to get experience working on different types of analysis. I noticed the investment in people and I feel proud of the work I do.”

Alexis Fernquest, Data Scientist at the Office for National Statistics.

To address these priority needs, mechanisms including enrichment and fellowship schemes, capability-building programmes and informal/ peer-to-peer mechanisms such as online courses, meet-ups and forums can prove effective. The following models are illustrations of how they can work in practice.

Mechanism

Enrichment and fellowship schemes

Model

Alan Turing Institute, Enrichment Scheme and Data Study Groups

One of the major goals of the Alan Turing Institute is to train new generations of data science and artificial intelligence leaders with the necessary breadth and depth of technical and ethical skills to match the UK's growing industrial and societal needs. The Enrichment Scheme offers students currently enrolled on a doctoral programme at a UK university the opportunity to join its research body. Doctoral students typically in their second or third year of study can undertake a 6, 9 or 12 month placement at the Institute's headquarters in London. Joining a community of more than 400 senior academics, early career researchers and PhD students, enrichment students have the opportunity to boost their skills and experience, enrich their research and make new collaborations during their time at the Alan Turing Institute.

The Data Study Groups are five-day 'collaborative hackathons', which bring together organisations from industry, government and the third sector, with talented multi-disciplinary researchers from academia. At each event, several

organisations bring their real-world problems to be tackled by small groups of highly talented, carefully selected researchers, with a diversity of thought. Researchers brainstorm and engineer data science solutions, presenting their work at the end of the week. Organisations get to quickly prototype possible solutions to their data science challenges, and researchers get an opportunity to put knowledge into practice and go beyond individual fields of research to solve real-world problems. Knowledge is exchanged among groups, and participants from both academia and the organisations posing challenges rapidly learn new skills during the week – from how to work in secure analysis environments to learning new data science methods and techniques, and tools for doing data science collaboratively in groups.

Mechanism

Graduate development programmes.

Model

Office for National Statistics advanced training

For the more advanced level training (eg PhD level), partnerships were made between the ONS, the Alan Turing Institute and several universities. The ONS has built relationships with many departments of the UK government which have set up Data Science 'hubs', allowing them to regularly exchange and communicate on the needs and skills needed. This should improve the flow of data science talent across the UK.

Model

Faculty Fellowship

The Faculty Fellowship (formerly ASI Data Science Fellowship) exists to ensure that the brightest academics get a chance to immerse themselves in working life, learn about artificial intelligence (AI) in business and help build the future of operational AI.

Since its founding in 2014, Faculty has trained and transitioned over 250 PhD STEM graduates into data scientist roles in industry. Taking place three times a year (January, May and September), the Fellowship is highly competitive and receives applications from 5 to 10% of the UK's physics, mathematics and engineering postgraduate research students. In part, this is because alumni go on to work for big names like Google, DeepMind, Facebook and Deliveroo.

Faculty believes that AI is the most important technology of our age, but that it is only valuable when applied in the real world – enhancing products, improving services, and saving lives. To apply AI, organisations need the right strategy, software and skills. This is why fellows are given the opportunity to build their AI skills while working on tangible, real-world problems during a six-week placement.

Model

Pivigo data science training

Pivigo is a data science marketplace and training company based in London. It helps organisations to innovate through data science by connecting them with their own community of data scientists. Its Science to Data Science programme trains and graduates some of the world's top scientific PhD talent in data science, with three programmes each year. Pivigo runs a five-week programme at its London campus or online. It works with large multinationals, charities, SMEs, and start-ups to help learners gain practical experience with data science technologies and technical skills in a commercial environment. The scheme offers students the opportunity to boost their skills, grow their networks and work alongside researchers.

Mechanism: Capability-building programmes

Model

The Cross-Government Data Science Accelerator

The Data Science Accelerator is a 12 week skills-building programme that gives analysts and aspiring data scientists from across the public sector, including central and local government, the opportunity to develop their data science skills. Created in 2015, this award-winning programme has been recognised for its impact on increasing data capability across the Civil Service. More than 150 participants from across the country have delivered a variety of projects, many of which have made a substantial difference to their public sector organisations⁴⁴.

44. Government Digital Service, Government Office for Science and Office for National Statistics. 2019. Introduction to the Data Science Accelerator programme. See <https://www.gov.uk/government/publications/data-science-accelerator-programme/introduction-to-the-data-science-accelerator> (accessed 15 April 2019).

“The exchange of ideas of an applied nature via a dialogue between experienced academics and practitioners has been key to the success of the learning program.”

Shakeel Khan,
Data Science Capability
Building Manager at
HMRC

Projects range from a ‘geospatial risk and impact’ tool designed to help local government strategically focus its service delivery to automating the process of categorising businesses using large sets of free text descriptions through machine learning algorithms⁴⁵.

Throughout the programme, participants are paired with a mentor who coaches them through the data challenges – and opportunities – that their projects bring. The programme is open for application four times a year and is free. Participants dedicate at least one day per week, usually attending one of the Accelerator’s hubs in London, Manchester, Newcastle, Newport, Sheffield or Taunton. Examples of previous projects can be found in the project directory and the Data Science Accelerator blog⁴⁶.

Model

Internal governmental initiatives

Some government departments have their own internal initiatives to upskill their workforces. For example, HMRC has a capability-building initiative that gives analysts and data scientists from across the public sector the opportunity to develop their data science skills. HMRC has also created a specific job role ‘Data Science Capability Building Manager’ to learn and disseminate applied data science knowledge. The collaborative work has been particularly beneficial for HMRC, lending methods from financial services that as a sector has

common objectives to manage customer relationships and mitigate non-compliance in a high volume financial transaction environment. Its masterclass initiative has led to further collaborative opportunities for thought leadership in data science and to hold data science problem-based forums. The masterclass courses have been delivered by academics from Edinburgh, and Imperial College London and delegates have attended from many UK government departments as well as other tax authorities across Europe.

Model

Prosperity Partnerships

Rolls-Royce is leading a large UKRI-EPSRC ‘Prosperity Partnership’ in Advanced Simulation and Modelling of Virtual Systems (ASiMoV), which is helping train the data scientists needed to process and analyse the petabytes of data generated by ultra-high fidelity exascale simulations. The programme combines leading edge research with PhD training and opportunity to work on business led challenges. By combining several of their highly regarded University Technology Centres with other leading UK universities and SME companies, Rolls-Royce is developing the skills needed for the integration of data, computational and physical sciences across a wide range of application areas.

45. Roberts S. 2018 Data Team Wins Future Policy Network Award. *Government Digital Service blog*. 19 February 2018. See <https://gds.blog.gov.uk/2018/02/19/data-team-wins-future-policy-network-award/> (accessed 15 April 2019).

46. Government Digital Service, Government Office for Science and Office for National Statistics. 2019. Previous Data Science Accelerator projects. See <https://www.gov.uk/government/publications/data-science-accelerator-programme/previous-data-science-accelerator-projects> (accessed 15 April 2019).



Chapter five

AREA FOR ACTION:

Enabling movement
and sharing of talent

AREA FOR ACTION:

Enabling movement and sharing of talent

“I can see a transition from a lot of individual contributor work, especially in the ingesting of data (cleaning it, making it available to use in a downstream process, for example modelling), to distribution of labour by specialism, and some of that labour is computer driven and not human driven.”

Ilya Zheludev, Chief Data Officer for Jasmine22.

To perform at their best, scientists and engineers need the right environment in which to work, one based on freedom to collaborate and access to different kinds of resource. Typically, industry and academia offer different experiences as working environments. Industry tends to offer access to funding and to large amounts of data in real time. Academia tends to offer freedom to explore and easy routes to cross-fertilisation with other subjects. The best research may well be done using both kinds of resource, and without forcing a long-term choice of sector upon the most talented people. Nowhere is this more urgent than in data science, which is an area of explosive growth and rapid change that is attracting talent from around the globe.

There are a number of ways to make moving between sectors a natural part of the data science career path. This is important on a variety of timescales from career move to week-long study through a secondment/internship.

In particular, there is a need to address the problem of the ‘one-way door’ out of academia which makes it difficult for researchers to return after spending time in industry or government, and one enabler to this movement is the ability of researchers to keep publishing when they are outside of academia.

Universities and the public sector in particular need to work to create a culture that nurtures and retains data science talent. This can include building and supporting cross-disciplinary data science teams, potentially disrupting traditional organisational structures.

Big internet companies are adding to pressure on universities, which are already struggling to retain professors and other employees. For sectors facing particular challenges with retention, there is a need to offer more incentives and address common barriers. One factor impacting on retention is that UK universities are seeing more and more interest in their intellectual property from big tech companies⁴⁷.

“Large institutions are slowly starting to understand how they can build IP and then retain it and monetise it for their own needs. If the UK government is interested in furthering the narrative of commercialisation of IP, it is natural progression that individuals working in academia could be encouraged or could wish to participate actively in commercialisation of knowledge. I have definitely seen a trend where young people are interested in going into academia and being part of a university that allows them to retain IP of anything they uncover themselves. That builds frameworks around allowing those individuals to turn that into commercial operations. I think that is a great step forward.”

Ilya Zheludev, Chief Data Officer for Jasmine 22.

47. Carey S. 2016 How UK universities are leveraging their intellectual property to benefit from the tech boom. *Techworld*. 8 November 2016. See <https://www.techworld.com/data/how-uk-universities-are-leveraging-their-ip-benefit-from-tech-boom-3648861/> (accessed 15 April 2019).

NEEDS AND RECOMMENDED ACTIONS

Need: Enable movement through braided careers

Supporting different ways of building a career is important as data science develops and is valuable across disciplines – as the Royal Society’s work on research culture shows⁴⁸. When data scientists are in high demand, enabling them to work across sectors and roles is valuable.

“There is definitely space for the best researchers from academia in industry. If you can take a leave of absence, keeping your level in academia, to spend time in industry, you could enrich your own research with knowledge in the field. Similarly, if you are in a really innovative sector and you see that for disrupting innovation you do not have all you need in your current role, you could benefit from the cooperation, input and autonomy of academia.”

Milton Luaces, Senior Manager at Accenture – Applied Intelligence.

RECOMMENDED ACTION

Create and fund joint positions across academia and industry

Funding bodies such as UKRI could support positions for joint appointments for the UK’s most talented researchers, who would be strongly in favour of a joint approach, so that a pool of excellence can be fostered at the interface of academia, industry and government. Universities and funders should give urgent attention to enhancing mechanisms to accommodate outstanding industrial research leaders in machine learning within the academic sector. This academic leadership is critical to inspiring and training the next generation of research leaders.

Need: Recognising diverse outputs

Recognising diverse outputs is important to support these braided careers. This means that work done in universities can be valued by industry and vice-versa.

RECOMMENDED ACTION

Commercialise research

The ways that universities encourage and support researchers in commercialising research and building spin-outs can influence researchers’ abilities to hold joint appointments between industry and academia. Universities may wish to consider their strategies for research commercialisation and policies on intellectual property in order to build an environment that supports cross-sector roles more freely.

“One of the things that I love about being in the UK is that collaboration is very important. It is a smaller environment with a lot of smart people.”

Kerem Sozucecer, Chief Technology Officer and co-founder of DeepZen.

48. Royal Society. (no date) Research culture. See <https://royalsociety.org/topics-policy/projects/research-culture/> (accessed 15 April 2019).

RECOMMENDED ACTION**Recognise diverse research outputs**

Government departments and industry are likely to benefit when they enable data scientists in research roles to publish their work wherever possible; conversely, universities need to recognise the value of data science experience outside universities, developed in the private and public sectors. Alternative outputs could be recognised on academic CVs. Changes to the Research Excellence Framework that focus on institutions rather than individuals could allow universities to better recognise the contribution of data science to broader research output.

Need: Establishing a coherent approach to data policy

Skills needs should be linked to the government's data strategy to ensure a joined-up approach.

RECOMMENDED ACTION**Make skills a core part of the National Data Strategy**

Responsibility for data policy is distributed across DCMS, GDS, Cabinet Office and DfE, but DCMS leads on delivering the National Data Strategy. This Strategy should enable departments to work closely together on data skills, building a coherent approach to delivering a healthy data science skills landscape. This will be important for the wider adoption of artificial intelligence.

MODELS AND MECHANISMS

There are a number of ways to make moving between sectors a natural part of the data science career path. Some mechanisms exist to foster long-term collaborative and business engagement networks and close interdisciplinary links. These include centres for doctoral training, industry fellowships, data residencies and innovative approaches to commercialisation. Here are some examples of existing models which highlight how the mechanisms are working in practice.

Mechanism

Centres for doctoral training

Model

Centre for Doctoral Training in Data Science and AI, The University of Edinburgh, and beyond

In 2013, the Engineering and Physical Sciences Research Council (EPSRC) invested £500 million in 115 Centres for Doctoral Training (CDTs), matched by more than £450 million from business, universities and other stakeholders. The University of Edinburgh has hosted a CDT in Data Science since 2014.

The CDT has two types of studentships including one with a PhD project in collaboration with an industry partner. For both courses the first year provides Masters level training in the core areas of data science along with a significant project. In years 2 – 4, students carry out PhD research in data science. In 2017, the UK's Science and Technology Facilities Council announced £10 million to train the next generation through supporting eight new CDTs in data intensive science. The centres include industrial partners and will offer comprehensive training in data intensive science through cutting edge research projects and a targeted academic training programme.

Model

DISCnet

The Data Intensive Science Centre is an STFC Centre for Doctoral Training providing a platform to train a new generation of data intensive scientists. The innovative education, training and research is delivered by a consortium of five universities from the South East Physics Network (SEPnet). The centre trains postgraduate students via world-leading research projects in particle physics and astrophysics and explores the untapped potential of these big data skills in diverse applications across a spectrum of industries. DISCnet currently has 70 non-academic partners.

Mechanism

Industry fellowships

Model

EPSRC Research Software Engineer Fellowships

The Research Software Engineering Fellowship is awarded to exceptional individuals in the software field, who demonstrate leadership and have combined expertise in programming and a solid knowledge of the research environment. As well as having expertise in computational software development and engineering, RSE Fellows should be ambassadors for the research software community and have the potential to be a future research leader in the RSE community.

“Close university and business collaboration resulted in more capability building, expansion of R&D activities of major industry – bringing new jobs and getting projects live faster.”

Professor Peter Buneman
FRS, University of Edinburgh

Model**UKRI Future Leaders Fellowships**

The Future Leaders Fellowship (FLF) is UKRI's flagship talent scheme, which aims to develop the next generation of research and innovation leaders in the UK. It will recruit and retain rising stars by attracting the brightest and best from at home and across the world. The FLF scheme will provide long-term funding for each Fellow (up to £1.2 million over an initial four years, with an option to extend to seven), allowing them to tackle difficult and novel challenges. A total of 550 FLFs will be awarded from 2019/20 to 2021/22 across six separate rounds, marking a significant investment to grow the UK's research and innovation base. Although the FLF scheme is not prescriptive of the research and innovation areas it supports, and Fellowships will be awarded on a competitive basis, there are several features that should make it attractive to those working in data science and AI. Whereas most existing Fellowship schemes fund only academic researchers, FLF also supports individuals in industry, as well as those working at the interface of academia, industry and the public sector, encouraging a new paradigm in career path that is mobile across all three. Operating across the breadth of UKRI will allow Fellows to take the most cross-cutting and interdisciplinary approaches to research and innovation. The open remit of the call allows for Fellowships to be held across a spectrum – from those with a background in AI wishing to apply their skills to a wide range of disciplines and challenges, to those who are from different disciplinary backgrounds, where AI could make a transformational contribution to that discipline or where that discipline could be brought to bear on the development of AI.

Model**Royal Society Entrepreneur in Residence scheme**

The Royal Society Entrepreneur in Residence (EiR) scheme aims to increase the knowledge and awareness in UK universities of cutting-edge industrial science, research and innovation⁴⁹. The scheme provides opportunities for outstanding industrial scientists and entrepreneurs to spend time working in a university to expose university staff and students to state-of-the-art industrial research and development, and the scientific challenges faced by industry. The scheme also allows universities to gain from expert advice aimed at promoting innovation and the translation of research by universities, and build confidence and understanding of business and entrepreneurship among staff and students.

Model**Royal Society Industry Fellowship scheme**

The Royal Society Industry Fellowship is a paid secondment scheme for academic scientists who want to work on a collaborative project with industry and for scientists in industry who want to work on a collaborative project with an academic organisation⁵⁰. Providing a basic salary for the researcher and a contribution towards research costs, the Fellowship aims to enhance knowledge transfer in science and technology between those in industry and those in academia in the UK. The scheme supports researcher-mobility and has run for over 30 years, bridging industry and academia for hundreds of scientists.

49. Royal Society. (no date) Entrepreneur in Residence. See <https://royalsociety.org/grants-schemes-awards/grants/entrepreneur-in-residence/> (accessed 15 April 2019).

50. Royal Society. (no date) Industry Fellowships. See <https://royalsociety.org/grants-schemes-awards/grants/industry-fellowship/> See (accessed 15 April 2019).

Model**Royal Commission for the Exhibition of 1851**

The Royal Commission for the Exhibition of 1851 awards three-year research fellowships to early career postdoc scientists or engineers of exceptional promise⁵¹. The Fellowship, which was founded in 1891 and has initiated the careers of thirteen Nobel laureates to date, is open to all nationalities and fields of science, including physical or biological sciences, mathematics, applied science, and any branch of engineering. The Commission also awards Industrial Fellowships to encourage industry – academia collaboration at doctoral research level, Industrial Design Studentships for postgraduates and, in partnership with the Royal Academy of Engineering, graduate Enterprise Fellowships for entrepreneurs.

Mechanism**Residencies****Model****Microsoft AI Residency program (US/UK)**

The Microsoft AI Residency program is a 12-month role designed to advance a career in machine learning research and engineering. The goal of the AI Residency is to help residents become creative and productive AI researchers, scientists and engineers. Residencies are open to BSc, MSc, and PhD graduates with substantial coursework in, but not limited to: computer science, electrical engineering, data science, mathematics, physics, economics, human – computer interaction, and computational biology.

Model**Uber AI Residency (US)**

Established in 2018, the Uber AI Residency is a 12-month training programme for recent college and Master's graduates, professionals who are looking to reinforce their AI skills, and those with quantitative skills and interest in becoming an AI researcher at Uber AI Labs or Uber Advanced Technologies Group (ATG). Uber AI Residents have the opportunity to pursue interests across academic and applied research. Uber is committed to an open and inclusive research mission that benefits the community at large, including contributing papers to top conferences and taking part in open-source projects.

Mechanism**Approaches to commercialisation****Model****IP free zone, Department of Computer Science and Technology, the University of Cambridge and beyond**

The IP free zone at the University of Cambridge is part of a more general framework set up by the former Head of Department of Computer Science and Technology, Professor Andy Hopper FRS. The strategy has been to minimise barriers to the formation of new companies while aligning incentives for staff and students, avoiding IP issues, providing mentoring, being helpful in every possible way, and not picking winners. Furthermore, this has been a cradle-to-grave approach ranging from undergraduate lectures to the maintenance of an industrial business club beyond the department. A total of 270 companies have been formed by staff and students (including Raspberry Pi), of which 50% are active with revenues of \$1 billion, and 18% sold for over \$40 billion.

51. Royal Commission for the Exhibition of 1851 (no date). Our awards. See <https://www.royalcommission1851.org/> (accessed 15 April 2019).

In the US, Carnegie Mellon and the University of Washington are currently working on a set of recommendations for commercial companies meant to provide a way for universities and companies to share talent more equally.

Model

UKRI Impact Acceleration Funding

Impact Acceleration Accounts (IAAs) are strategic awards provided to research organisations to support knowledge exchange and accelerate the impact of research. IAAs allow organisations to respond in more flexible, responsive and creative ways appropriate to their strategic priorities, enabling impact to be achieved in an effective and timely manner, for example, through secondments and exchanges, user engagement, proof of concept, and by building capacity for work across disciplines.



Chapter six

AREA FOR ACTION:

Widening access to data
in a well-governed way

AREA FOR ACTION:

Widening access to data in a well-governed way

“The more sensitive the industries are, the more difficult the partnership with academia. There are sectors that are quite sensitive on data: for instance, financial services, pharmaceuticals and government. One challenge is that usually they want people working on premises, which is difficult for academics.”

Milton Luaces, Senior
Manager at Accenture –
Applied Intelligence.

Access to good data can ensure that data scientists get necessary experience with ‘real world’ problems that is so important in data science. But more importantly, this will enable the use of data science skills for public and commercial benefit. However, it is important that this is done in an ethical and well-governed way.

Previous work by the Royal Society and the British Academy in Data Management and Use: Governance for the 21st Century, argues for a principled approach to the governance of data use. Central to this is that we should govern the use of data in such a way that makes human flourishing central, with data used to benefit people and communities. Data scientists should aim to open up data for social good and do so in a well-governed way.

Ways to create a professional ethos for data scientists, to share good practice in use of data and to agree shared codes for ethical collection and use of data will be important areas to explore further to ensure that data can be used procure diverse benefits for society. Professional bodies such as the Royal Statistical Society, with its data science section, and the British Computer Society are already exploring options for accreditation of data science courses to demonstrate the quality of course content. It will also be valuable to explore the development of shared codes of conduct or practice for data science to raise awareness of the need for an ethical and well-governed approach to data collection and use and to ensure that a concern with ethics is appropriately reflected in course content. Such professional bodies – working with the Centre for Data Ethics and Innovation, the Ada Lovelace Institute, the Open Data Institute and the many other groups promoting ethical use of data – can help explore and establish a positive professional ethos and ethics for data science.

“Making [data] available between academia and industry is really hard, because a lot of the time data is really sensitive, and it can also have personal information in there. Mechanisms that enable that sharing, in a legal, compliant way are really important. How can someone push the boundaries of knowledge in a domain if they cannot work from real data in that domain? I would say that governments can have quite a lot of involvement here. It is about collaboration and being able to work problem statements that are based on proprietary data.”

Ilya Zheludev, Chief Data Officer for
Jasmine 22.

NEEDS AND RECOMMENDED ACTIONS

Need: Opening up data and providing secure access.

To get the best value from data for the widest range of organisations means that opening data in a secure and well-governed way should enable societal benefit to be accessed most easily.

RECOMMENDED ACTION

Encourage data sharing where possible

Greater transparency of private sector data could help build public trust in the use of data and how their data is used for decision-making purposes. The public sector could usefully consider how to widen access to its data, including sharing data, and data challenges to researchers. Journal editors should normally ensure that data are being made available to other researchers in their original form, or via appropriate summary statistics where sensitive personal information is involved.

Standardisation of open data can enable the sharing of data and challenges, providing an amenable data environment for researchers. However, the majority of data held by organisations is not open, and unlikely to be so due to personal disclosure and commercial restrictions. In areas where there are datasets unsuitable for general release, further progress in supporting access to public sector data could be driven by creating policy frameworks or agreements which make data available to specific users under clear and binding legal constraints to safeguard their use, and set out acceptable uses. Government should further consider the form and function of such new models of data sharing.

The public sector needs to work out how to widen access to give university researchers access to better public data. Continuing to ensure that data generated by charity-funded and publicly funded research are open by default will be critical in supporting wider uses of research data. Journals should normally insist, as a condition of publication, on data being made available to other researchers in their original form, or via appropriate summary statistics where personal information is involved. Of course such a policy has additional potential benefits of enhancing reproducibility in research and increasing transparency of decision making, in both the public and private sectors.

The Royal Society has recently published the report, *Protecting privacy in practice*, which sets out how use of government data could be enabled by Privacy Enchanting Technologies (PETs).

RECOMMENDED ACTION

Donate data science talent

There is value in enabling data scientists to donate their time to applying data science to societal challenges. For example, through *probono* project work along the lines of DataKind UK, RSS Statisticians for Society and hackathons.

“We have amazing data in the NHS and that is definitely a resource that is worth staying for.”

Dr Amy Nelson, Senior Research Associate at UCL Institute of Neurology and a junior doctor.

Need: Providing the computing power for use by the growing data science community.

As the data science community grows there will be a need for greater access to high power computing, and to GPUs for artificial intelligence and machine learning, so that data scientists can realise their potential.

RECOMMENDED ACTION**Provide access to computing power**

Improving the UK's computing research infrastructure will better enable data scientists to access the necessary computing power to release the value from data and address research challenges, and will be vital for the UK to remain competitive with other countries such as the US and China. BEIS and UKRI could usefully consider the need for continuing to improve access for data scientists working across all disciplines to high-power computing, and this could helpfully be included as part of the UKRI Infrastructure Roadmap.

MODELS AND MECHANISMS

Mechanisms such as collaborative events and partnerships, data stores and APIs, offices of data analytics and data centres/institutes are important ways to bring data scientists and data together and the models below show how this can work, even for organisations that do not have the resources to hire data scientists themselves.

Mechanism

Collaborative events and partnerships

Model

Charity DataDives, DataKind UK

A hackathon brings together a range of people to generate an outcome, usually software projects. They can be associated with computer programmers, software developers, data scientists and, often, subject-matter-experts. The charity DataKind supports charities and social enterprises large and small across a variety of issue areas. It runs hackathon events called 'DataDives' where charities and social enterprises work alongside teams of volunteer data scientists, analysts, developers and designers using data to gain insight into their programmes and to increase their impact.

Hackathon style events can stimulate more engagement between the academic community and social projects as there are lots of skills within universities that are both expensive and in short supply within the third sector.

DataKind also run longer-term engagements over 6 – 9 months to build a data science solution (DataCorps) and have monthly office hours which any non-profit or social change organisation can sign up to for advice.

Model

Royal Statistical Society – Statisticians for Society

The Statisticians for Society initiative was launched in 2014, to help statisticians offer their skills to charities and other socially useful initiatives that need their professional expertise⁵². Many third sector organisations are keen to explore the use of data for decision making and service improvements. There is a growing need for them to provide evidence of their impact, but due to lack of capacity and appropriate skills needed for data analysis, some charities are unable to fully demonstrate the value of their work. As one of the leading voices for promoting the importance of data and evidence, the Royal Statistical Society supports statisticians in helping charities in making a difference. Volunteers can provide the tools and guidance for undertaking data analysis. Statisticians collect, analyse and interpret data across a wide range of industries and topics; they are skilled at designing methods for collecting data and regularly tasked with analysing data to spot patterns and trends; and they can manipulate data to identify relationships and make future predictions. Following this, they produce reports and summaries that communicate their findings.

52. Royal Statistical Society. 2019 Statisticians for Society. See http://www.rss.org.uk/RSS/Get_involved/Statisticians_for_Society/RSS/Get_involved/Statisticians_for_Society.aspx?hkey=c7977c58-1558-495a-9e5a-e99d64ea9cfd (accessed 15 April 2019).

Mechanism**Data stores****Model****The London Datastore**

The London Datastore is a free and open data-sharing portal where anyone can access data relating to the capital⁵³. It is one of the Greater London Authority's (GLA) flagship projects and is a platform through which many of the Smart London Plan objectives are delivered. Researchers are encouraged to visualise or build apps from the data available on the site.

Model**ONS API**

The Office for National Statistics API makes datasets and other data available programmatically, allowing researchers to filter datasets and directly access specific data points.

Model**UK Data Service**

The UK Data Service is the UK's only nationally funded research infrastructure for the curation and provision of access to social science data and its practices, especially around secure access to data and data curation, have been influential across the world. Funded by the Economic and Social Research Council (ESRC) to meet the data needs of researchers, students and teachers from all sectors, its unique collection of social science data resources includes major UK government-sponsored surveys, cross-national surveys, longitudinal studies, UK census data, international aggregate, business data, and qualitative data. It brings together several important past investments including the Economic and Social Data Service, Question Bank, Qualidata and Census Programme.

Climate change, ageing, security threats, the provision of better public services and a more productive society all call for policies and solutions that are informed by evidence-based research and innovation. Access and training services through the UK Data Service enable impactful research to influence national and regional policy and develop research excellence across all sectors. Its excellence in collecting, storing, analysing and sharing collections of complex data has helped it build a trusted reputation and become a critical part of the UK's research infrastructure.

53. London Datastore 2019. Welcome to the Datastore. See <https://data.london.gov.uk/> (accessed 15 April 2019).

BOX 2

Open Data Skills Framework
(The ODI)

Open data is a relatively new field. Its potential is being realised increasingly as it is slowly integrated into the strategies of organisations. Those working with open data or on open data initiatives often have to learn the skills as they go. There is no clear language to describe the knowledge and skills of those working with open data; nothing with which to benchmark a single individual's expertise, and point to where they are in their learning journey.

This is why the Open Data Institute has created the open data skills framework, a simple, three-tier framework that describes the knowledge and skills of anyone interacting with open data, from beginner through to expert level⁵⁶.

The skills framework enables learners to identify where they are in their learning journey. Everybody starts their journey as an 'explorer'. They may then wish to focus on either strategic or practical skills, or both. Ideally, every learner will eventually apply their learning back to their sector to help solve sector-specific challenges, and drive change in their domain.

Mechanism

Data stores

Model

Transport for London API

Transport for London (TfL) is the local public body responsible for public transport in London⁵⁴. Every year, TfL ensures the transportation of 1.37 billion people, with a network length of 402 km, which is equivalent to 83.6 million km travelled per year⁵⁵. Over the past ten years TfL has made a significant amount of data accessible to the public free of charge, including timetables, service status and disruption information. This has allowed the market to develop exponentially with the introduction of new products and services. TfL is now considered as a leader in publishing open data through APIs, the Cloud, the internet and across its physical network. It has created over 700 jobs and brought £14 million per year in GVA, enabling development of UK's skills in data.

Mechanism

Offices of Data Analytics

Model

Nesta's programme of Offices
of Data Analytics

The Offices of Data Analytics (ODA) programme helps cities and regions join up, analyse and act upon data from multiple sources to reform public services⁵⁷. As of December 2018, there were nine initiatives that classify as Offices of Data Analytics across the UK. The model allows multiple organisations to join up, analyse and act upon data sourced from multiple public sector bodies to improve services and make better decisions.

54. Transport for London 2019. Facts & figures. See <https://tfl.gov.uk/corporate/about-tfl/what-we-do/london-underground/facts-and-figures> (accessed 15 April 2019).

55. Deloitte 2017. Assessing the value of TfL's open data and digital partnerships. See <http://content.tfl.gov.uk/deloitte-report-tfl-open-data.pdf> (accessed 15 April 2019).

56. Open Data Institute. 2016 Open Data Skills Framework. See <https://theodi.org/article/open-data-skills-framework/> (accessed 15 April 2019).

57. Eaton M and Bertocin C 2018. State of Offices of Data Analytics (ODA) in the UK. See: https://media.nesta.org.uk/documents/State_of_Offices_of_Data_Analytics_ODA_in_the_UK_WEB_v5.pdf (accessed 15 April 2019).

Mechanism:**Data Centres / Institutes****Model****Health Data Research UK (HDR UK)**

Health Data Research UK is uniting the UK's health data to make discoveries that improve people's lives. By bringing together the sharpest scientific minds, and providing safe and secure access to rich health data, it aims to better understand diseases and discover new ways to prevent, treat and cure them⁵⁸. Its vision is for large-scale data and advanced analytics to benefit every patient interaction, clinical trial, biomedical discovery and improve public health. To achieve this, HDR UK is leading an ambitious training and talent programme, and will create a cohort and network of thousands of health data scientists spanning all career stages, from school-leaver to senior research manager and international opinion leaders. The UK has a rich and diverse scientific talent base, thanks to the strength of the NHS, its academic institutions and innovative scientific and digital industries. HDR UK plans to harness this, bring on board international peers, to create an intelligent cohort of health data scientists that will dramatically change medical research, and open up new, faster, smarter pathways to patient care.

Model**National Innovation Centre for Data (NICD)**

The National Innovation Centre for Data (NICD) is a unique new facility that delivers data analytics skills into industry and the public sector by exploiting the knowledge and expertise currently locked within universities. A flexible rolling programme of collaborative projects focused on organisations' specific challenges and opportunities will transfer practical data skills into the workforce of those organisations. These projects will be supported by a range of related activities, including awareness-raising events, themed business and technical seminars and technical training courses. As a result of engagement with the Centre, organisations will be able to increase their productivity by optimising their existing operations, and to grow by launching new data-driven products and services.

58. Medical Research Council. 2019. Health Data Research Institute. See <https://mrc.ukri.org/about/institutes-units-centres/uk-institute-for-health-and-biomedical-informatics-research/> (accessed 15 April 2019).



Conclusion

Conclusion

Data professionals are in high demand from employers. Over the last five-and-a-half years, there has been a sharp rise in UK job-listings for 'Data Scientists and Advanced Analysts' (+231%) driven predominately by increased numbers of vacancies for Data Scientists (+1287%) and Data Engineers (+452%)

This report has focused on Data Scientists and Advanced Analysts at the top end of analytical rigour because this is where demand has grown the most. However, the data shows interesting results for data professionals across the spectrum. Moreover, our findings are likely to underestimate the demand for data skills as many jobs are not advertised online. Further analysis is needed to quantify the number of employed workers per opening.

There is a clear need for collaborative, sustainable mechanisms to develop data talent in academia, and the charity, private and public sectors, and to allow data scientists to move across these sectors. This report promotes

a vision for the sharing of data science talent across all sectors. By identifying four major areas of action with recommendations for addressing priority needs across the data science talent pipeline, from school to advanced professionals, we are hopeful that we can achieve our vision for the UK to be a leading data science research nation with a sustainable flow of expertise and a healthy data science skills landscape.

This report also sets out a wide variety of existing models and mechanisms that could be used more widely, from fellowship schemes to data stores, that represent good practice or innovation in supporting data science career development and mobility. The report features several contributions from data scientists from a range of backgrounds, organisations and roles sharing their career experiences: these are also available as a separate publication, *Dynamics of Data Scientists: what data professionals say about data science*.

Appendices

Image

Caption goes here.

APPENDIX 1:

Acknowledgements

Thank you to Will Markow and Jonathan Coutinho at Burning Glass Technologies for providing the labour market data and advice with the analysis, the project steering group for providing advice and guidance, the reviewers and all of those involved in the workshops and case study interviews.

Working Group members

The members of the Working Group involved in this report are listed below. Members acted in an individual and not a representative capacity, and declared any potential conflicts of interest. Members contributed to the project on the basis of their own expertise and good judgement.

Working Group members

Professor Andrew Blake FREng FRS, AI consultant and Chairman of Samsung's AI Research Centre in Cambridge UK (Chair)

Professor Peter Buneman MBE, FRSE, FRS, Professor of Database Systems, School of Informatics, University of Edinburgh

Sherry Coutu CBE, serial entrepreneur, former CEO, angel investor and non-executive director

Matthew Fryer, VP, Chief Data Science Officer, Hotels.com

Dr Robert Hercock, Chief Research Scientist in the British Telecommunications Security Research Practice

Professor Marta Kwiatkowska FRS, Department of Computer Science, Oxford University and Professorial Fellow of Trinity College

Professor Emma McCoy FRSS FIMA, Vice-Dean (Education), Faculty of Natural Science, Imperial College London

Dr Tom Smith, Managing Director of the Data Science Campus, Office for National Statistics

Royal Society staff

Royal Society Secretariat

Dr Natasha McCarthy, Head of Policy (Data)

Jennifer Panting, Policy Adviser and Project Lead

Staff from across the Royal Society contributed to the development of the project

Dr Frances Bird, Policy Adviser

Connie Burdge, Project Coordinator

Dr Claire Craig CBE, Chief Science Policy Officer

Dr Frank Fourniol, Policy Adviser

Omar Jamshed, Press Officer

David Montagu, Policy Adviser

Karen Newman, Design Manager

Selina Patel, Policy Intern

Dr Mahlet Zimeta, Senior Policy Adviser

Previous Royal Society staff who contributed to the development of the project

Lara Gardellini, Programme Coordinator

Louise Pakseresht, Senior Policy Adviser

Reviewers

This report has been reviewed by an independent panel of experts. The Review Panel members were not asked to endorse the conclusions or recommendations of the report, but to act as independent referees of its technical content and presentation. The Royal Society gratefully acknowledges the contribution of the reviewers.

Reviewers

Professor Dame Wendy Hall, DBE, FREng, FRS, Regius Professor of Computer Science at the University of Southampton, UK

Professor Frank Kelly CBE, FRS, Professor of the Mathematics of Systems at the Statistical Laboratory, University of Cambridge, UK

Professor Tom McLeish FRSC, FRS, Professor of Natural Philosophy, University of York

Dame Jil Matheson DCB FAcSS, former National Statistician of the United Kingdom

Hetan Shah, Executive Director of the Royal Statistical Society

Sir Bernard Silverman FRS, former Chief Scientific Adviser at the Home Office

Expert reader

Dr Jonathan Shaw, Technical Specialist in the Economics Department at the Financial Conduct Authority and Research Fellow at the Institute for Fiscal Studies

Interview and case study participants

The Royal Society gratefully acknowledges the contribution of the case study participants.

Interview and case study participants
Aimee, Government Communications Headquarters (GCHQ)
Professor Graham Cormode, Professor in Computer Science at the University of Warwick
Alexis Fernquest, Data Scientist at the Office for National Statistics (ONS), and former Data Analytics Apprentice
Dr James Hetherington, Director of Research Engineering at the Alan Turing Institute
Professor Frank Kelly FRS, Emeritus Professor of the Mathematics of Systems at the University of Cambridge
Kevin Koene, former Junior Data Scientist at The One Campaign
Milton Luaces, Senior Manager at Accenture – Applied Intelligence
Nick Manton, Head of Data Science at the Government Digital Service, and cross government Head of Community for Data Scientists within the DDaT (Digital Data and Technology) profession.
Dr Amy Nelson, Senior Research Associate at UCL Institute of Neurology and a junior doctor
Chanuki Illushka Seresinha, Lead Data Scientist at Popsa, a start-up using AI to automatically curate photo content into designed physical products
Dr Tom Smith, Director of the Data Science Campus at the Office for National Statistics
Kerem Sozuceker, Chief Technology Officer and co-founder of DeepZen, a start-up which specializes in creating emotional and expressive human voice with AI
Dr Damon Wischik, Lecturer in the University of Cambridge's Computer Laboratory and former recipient of a University Research Fellowship from the Royal Society
Dr Maria Wolters, Reader in Design Informatics at the University of Edinburgh
Ilya Zheludev, Chief Data Officer for Jasmine22, a new technology venture being built by HSBC based in Hong Kong

Workshop participants

The Royal Society would like to thank all those who contributed to the development of this project through submission of evidence and attendance at the following workshops.

March 2018: *What's different about data science? scoping roundtable*

This roundtable asked 'what's different about data science?' to gather evidence on attracting and cultivating data science talent, skills and the models and mechanisms to enable a thriving landscape. Guests discussed whether the UK has the skills capacity to deliver the potential of data science and AI and how to ensure a healthy landscape that enables talent to grow and flow between sectors.

May 2018: *Contextualising the disruption evidence gathering workshop*

The aim of this workshop was to determine patterns in the drivers for the movement of data scientists and interrogate a number of previously identified models for upskilling, retaining and sharing data science talent.

September 2018: *Data talent in Newcastle*

This workshop was to discuss the demand for data professionals across Newcastle and North East England (in all sectors – public, industry and academia), how skills gaps are currently being met and what more could be done to develop a pipeline of talent. Guests represented local universities, training centres and businesses in data/analytics roles or the recruitment/ training of data professionals.

January 2019: *Reviewing the emerging policy recommendations*

A lunch discussion was held at the Royal Society to discuss the outcomes of the project on the UK data science workforce with key stakeholders. Guests were asked to review the project findings and suggest ways to improve, refine and revise the content, particularly the areas of action and recommendations for change.

March 2019: *Data talent in Wales*

This workshop was to discuss the demand for data professionals across Wales (in all sectors – public, industry and academia), how skills gaps are currently being met and what more could be done to develop a pipeline of talent.

APPENDIX 2:

Glossary

Algorithm: A set of rules a computer follows to solve a problem.

Artificial intelligence (AI): An umbrella term for the science of making machines smart.

API: Application Programming Interface.

BEIS: The Department for Business, Energy and Industrial Strategy.

Big data: Large and heterogeneous forms of data that have been collected without strict experimental design. Big data is becoming more common due to the proliferation of digital storage, the greater ease of acquisition of data (eg through mobile phones) and the higher degree of interconnection between our devices (ie the internet).

Data: Numbers, characters or images that designate an attribute of a phenomenon.

DCMS: The Department for Digital, Culture, Media and Sport.

DfE: The Department for Education.

DSAA: Classification code used by Burning Glass Technologies to group together Data Science and Advanced Analytics job vacancies.

DSA: Classification code used by Burning Glass Technologies to group together Data Science and Analytics job vacancies.

GDS: The Government Digital Service.

Hadoop: An open source framework that manages data processing and storage for big data applications.

Java: A programming language.

Machine learning: A set of rules that allows systems to learn directly from examples, data and experience.

Metadata: 'Data about data', contains information about a dataset. For example, this information could include why and how the original data was generated, who created it and when. It may also be technical, describing the original data's structure, licensing terms, and the standards to which it conforms.

MOOC: A Massive Open Online Course (MOOC) is an online course aimed at unlimited participation and open access via the web.

Python: A programming language.

UKRI: UK Research and Innovation (UKRI) is the national funding agency investing in science and research in the UK.

R: A programming language.

SAS: A software suite developed by the SAS Institute for analytics purposes.

Scala: A programming language.

SIC-1: The Standard Industrial Classification (SIC) is a system for classifying industries according to their economic activity.

STEM: A term used to group together Science, Technology, Engineering and Mathematics.

APPENDIX 3:

Data

This appendix includes additional data tables, explanations about the Burning Glass Technologies methodology and other findings from the data.

Understanding the Burning Glass**Technologies methodology: Skills clusters**

Skills clusters are groupings of related skills. Burning Glass Technologies has developed a taxonomy of more than 500 skills clusters by grouping skills that often travel together in job postings. Clustering has been focused on the most frequently occurring skills across a range of industries, in addition to skills in emerging areas. Burning Glass Technologies labour market analysts used the following three criteria to group skills:

- related skills eg the skill cluster ‘Statistical Software’ includes skills such as R, SAS, and SPSS;
- skills that travel together eg the skill cluster ‘Administrative Support’, includes skills such as meeting planning/facilitation, calendar management, travel arrangements, and appointment setting; and
- skills that are trained together eg the skill cluster ‘Lean Manufacturing’ includes skills such as Kanban, Kaizen, Six Sigma, and Lean Six Sigma.

Skills are grouped for ease of analysis of broader talent requirements. Burning Glass Technologies skills hierarchy and grouping has been created using a combination of hierarchical clustering algorithms and assessment of skills similarity based on job postings. Using a variety of distance metrics including Cosine, Dice, Jaccard, and others, the similarity of all skills combinations is determined. Depending on how close two skills cluster are together based on the similarity measures, they get assigned to the same skills cluster. The final step included manual reviews of the clusters to resolve any unclear cases.

Missing salary information

The number of postings and the average salary have been calculated for the year 2013 and the last four full quarters (July 1, 2017 – June 30, 2018). However, a large proportion of postings have no salary attached so these results should be seen as indicative rather than definitive. Intriguingly the salary for the ‘Data Scientists’ occupation showed only a small increase, despite the huge increase in postings.

Numerous drivers might explain the salary/demand anomaly for data scientists. Perhaps employers are leaving some 'wiggle room' in negotiations. Perhaps the nature of the jobs has shifted as data science transitioned from a sophisticated role into one that leverages different skillsets. Perhaps employers are adjusting their requirements because of the supply and demand dynamics that are not in their favour – eg asking for a BA instead of a PhD. Perhaps the jobs are newer and so there is a limited supply of people in senior roles with higher paying salaries.

There has been a reduction in the level of experience requested since 2013, with nearly half of all Data Scientist postings requesting just 0 – 2 years of experience. Looking at skillsets within occupations would help to try to answer this question, but beyond the scope of this project.

Qualification levels

Qualification levels requested for the 'Data Scientists and Advanced Analysts' category have broadly increased. In 2013, 34% of such postings required Level 6 (first degree) or Level 7 (MSc or upwards) skills, but by 2017/18 this had increased to 42%. This was most acutely seen in the 'Data Scientist' occupation, where half of all postings required Level 6 or 7.

Certifications

Data science jobs are not as heavily certificated as other fields. This could be because there is no supply of industry recognised credentials at present. For example, in 2017 – 18 while 14.9% of 'Budget Analyst' roles mentioned relevant accounting certifications, only 1.1% of 'Data Engineer' roles mentioned the relevant Electrotechnical Card Scheme (ECS) certification. The IT field has more certification demand because the community has long-standing credentials, for example ITIL (a set of detailed practices for IT service management) and CompTIA (performance-based exams that certify foundational IT skills across a variety of devices and operating systems).

ADDITIONAL DATA TABLES

APPENDIX TABLE 1

The classification system used by Burning Glass Technologies to group data science and analytics jobs in order of increasing analytical rigour (left to right).

Data-Driven Decision Makers	Functional Analysts	Analytics Managers
<ul style="list-style-type: none"> • Chief Executive Officer • Chief Information Officer / Director of Information Technology • Compensation and Benefits Manager • Financial Manager • Human Resources Manager • IT Project Manager • Logistics Manager • Marketing Manager • Operations Manager • Procurement Manager • Product Manager • Quality Control Systems Manager • Talent Acquisition Manager 	<ul style="list-style-type: none"> • Actuary • Budget Analyst • Business / Management Analyst • Clinical Analyst / Clinical • Documentation and Improvement Specialist • Clinical Data Manager • Compensation / Benefits Analyst • Credit Analyst / Authoriser • E-Commerce Analyst • Financial Analyst • Financial Examiner / Auditor • Fraud Examiner / Analyst • Geographer / GIS Specialist • HRIS Analyst / Specialist • Human Resources Analyst • Logistics / Supply Chain Analyst • Market Research Analyst • Operations Analyst • Pricing Analyst • Researcher / Research Associate • Risk Analyst • Risk Consultant • Search Engine Optimisation Specialist • Security / Defence Intelligence Analyst • Social Science Researcher • Survey Researcher 	<ul style="list-style-type: none"> • Chief Executive Officer • Chief Information Officer / Director of Information Technology • Compensation and Benefits Manager • Director of Risk Management • Financial Manager • Human Resources Manager • IT Project Manager • Marketing Manager • Product Manager • Risk Manager • Talent Acquisition Manager

Data Systems Developers	Data Analysts	Data Scientists and Advanced Analysts	KEY Framework category
<ul style="list-style-type: none"> • Business Intelligence Architect / Developer • Computer Systems Engineer / Architect • Data Warehousing Specialist • Database Administrator • Database Architect • Hadoop Developer • Systems Analyst 	<ul style="list-style-type: none"> • Business Intelligence Analyst Data / Data Mining Analyst 	<ul style="list-style-type: none"> • Biostatistician • Data Engineer • Data Scientist • Economist • Financial Quantitative Analyst • Statistician 	<ul style="list-style-type: none"> • Occupation

APPENDIX TABLE 2

Regional data and Data Science and Advanced Analytics (DSAA) jobs.

Region/Nation	Data jobs 2013	Data jobs 2018	% increase	DSAA jobs 2013	DSAA jobs 2018	% increase
Northern Ireland	4,987	11,902	139%	46	305	563%
Scotland	36,388	41,117	13%	528	1,114	111%
Wales	9,198	14,507	58%	121	216	79%
England	654,242	837,307	25%	7,130	22,527	216%
East Midlands	30,957	37,743	22%	200	474	137%
East of England	53,295	61,725	16%	551	1,928	250%
Greater London	283,817	345,164	22%	4,131	14,066	240%
North East	7,444	11,095	49%	54	175	224%
North West	45,286	61,871	37%	320	1,182	269%
South East	112,721	126,551	12%	1,086	2,305	112%
South West	41,066	59,581	45%	276	903	227%
West Midlands	41,771	86,563	107%	256	748	192%
Yorkshire and the Humber	37,885	47,014	24%	256	746	191%
TOTAL	704815	904833	28%	7825	24162	209%

APPENDIX TABLE 3

The top 10 skills listed in DSAA job adverts (2013).

This table shows the skills which occurred the most in 2013. This is measured in terms of the proportion of Data Science and Advanced Analyst (DSAA) job adverts which specified the skill as a requirement for the role. There were a total of 682 skills included in this analysis. This table displays the top ten most frequently occurring skills.

Rank	Skill	Number of DSAA job adverts requiring this skill		Percentage of DSAA job adverts requiring this skill	
		2013*	2017 – 18**	2013	2017 – 18
1	Research	2,042	5,279	25%	20%
2	Communication Skills	1,843	4,849	23%	18%
3	Statistics	1,657	2,800	20%	10%
4	Economics	1,340	1,984	16%	7%
5	SAS	1,275	2,323	16%	9%
6	Microsoft Excel	1,126	1,585	14%	6%
7	SQL	1,048	7,226	13%	27%
8	Data Analysis	841	2,858	10%	11%
9	Statistical Analysis	811	1,630	10%	6%
10	C++	754	2,166	9%	8%

*Out of 8,157 job adverts included in this analysis. **Out of 27,033 job adverts included in this analysis.

APPENDIX TABLE 4

The top 10 skills clusters listed in DSAA job adverts (2013).

This table shows the skills clusters which occurred the most in 2013. This is measured in terms of the proportion of Data Science and Advanced Analyst (DSAA) job adverts which specified the skills clusters as a requirement for the role. There were a total of 278 skills clusters included in this analysis. This table displays the top ten most frequently occurring skills clusters.

Rank	Skills cluster	Number of DSAA job adverts requiring this skill cluster		Percentage of DSAA job adverts requiring this skill cluster	
		2013*	2017 – 18**	2013	2017 – 18
1	Statistics	1,908	7,406	23%	27%
2	Statistical Software	1,797	8,894	22%	33%
3	Data Analysis	1,628	10,636	20%	39%
4	Economics	1,532	5,086	19%	19%
5	SQL Databases and Programming	1,151	15,532	14%	57%
6	Data Science	1,056	25,042	13%	93%
7	Project Management	918	4,046	11%	15%
8	Medical Research	867	2,200	11%	8%
9	Scripting Languages	829	23,450	10%	87%
10	C and C++	757	4,344	9%	16%

*Out of 8,157 job adverts included in this analysis. **Out of 27,033 job adverts included in this analysis.



The Royal Society is a self-governing Fellowship of many of the world's most distinguished scientists drawn from all areas of science, engineering, and medicine. The Society's fundamental purpose, as it has been since its foundation in 1660, is to recognise, promote, and support excellence in science and to encourage the development and use of science for the benefit of humanity.

The Society's strategic priorities emphasise its commitment to the highest quality science, to curiosity-driven research, and to the development and use of science for the benefit of society. These priorities are:

- Promoting excellence in science
- Supporting international collaboration
- Demonstrating the importance of science to everyone

For further information

The Royal Society
6 – 9 Carlton House Terrace
London SW1Y 5AG

T +44 20 7451 2500

E science.policy@royalsociety.org

W royalsociety.org

Registered Charity No 207043



ISBN: 978-1-78252-395-6

Issued: May 2019 DES5847