

Dynamics of data science skills: What data professionals say about data science – career stories

We interviewed data professionals across government, academia and industry working in the broad field of data science, particularly those who have moved across sectors, to better understand the opportunities and challenges of their careers, and the drivers and blockers for movement.

Their career stories illustrate the richness of the data science landscape and the complex network of factors that impact on career choice. They discuss the mechanisms that can enable data professionals to thrive in multiple roles, or move between them, and highlight where skills gaps exist and how these can be filled.

Here's what data professionals have to say about their industry.

This publication accompanies the Royal Society report, [*Dynamics of data science skills: How can all sectors benefit from data science talent?*](#)

Contents

The future is start-ups Kerem Sozuceger – Chief Technology Officer and co-founder of DeepZen.	3
Revolutionising the NHS through data science Dr Amy Nelson – Senior Research Associate at UCL Institute of Neurology and a junior doctor.	5
Leading the data science profession within UK government 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.	7
Using data from the web to understand human behaviour Chanuki Illushka Seresinhe – Lead Data Scientist at Popsa and Visiting Researcher at the Alan Turing Institute.	10
Contributing to academic research through the creation of software Dr James Hetherington - Director of Research Engineering at the Alan Turing Institute	13
From Silicon Valley, US to Silicon Fen, UK Dr Damon Wischik – Lecturer in the University of Cambridge’s Computer Laboratory.	15
An analytics apprenticeship at the UK’s largest independent producer of official statistics Alexis Fernquest – Data Scientist at the Office for National Statistics (ONS).	17
What will transport within cities look like in twenty years? Understanding government data science challenges Frank Kelly CBE FRS – Emeritus Professor of the Mathematics of Systems at the University of Cambridge.	19
Applying data science to end extreme poverty Kevin Koene – Previously a Junior Data Scientist at The One Campaign.	20
The skills that define a data scientist Milton Luaces – Senior Manager at Accenture.	22
Applying cutting-edge techniques to impact on public good Tom Smith – Director of the Data Science Campus at the Office for National Statistics.	24
Fostering entrepreneurship and approaching commercialisation Ilya Zheludev – Chief Data Officer for Jasmine22.	26
The importance of continued learning Dr Maria Wolters – Alan Turing Fellow and Reader in Design Informatics at the University of Edinburgh.	29
Rethink UK university teaching to produce tech-ready graduates at scale Graham Cormode – Professor in Computer Science at the University of Warwick.	31
Entry-level Analyst to Data Scientist at Government Communications Headquarters (GCHQ) Aimee – Capability Researcher at GCHQ.	33

The future is start-ups

Kerem Sozuceger is Chief Technology Officer and co-founder of DeepZen, a start-up which specializes in creating emotional and expressive human voice with AI.



CAREER HISTORY

From school to start-ups to corporations, across USA, Turkey, and UK

Personally, I never liked school. I was always about graduating and paying my dues, meaning having a college graduate degree, but I knew I had to start working to have an impact. I graduated a semester early even though I took half a year off, because I wanted to get out into the real world and start working.

I do believe in the education system, school and university, and that it makes you a better person, and a person who is better at planning. But you are not really creative, you are just learning a bunch of stuff and a lot of it isn't useful knowledge. It is not the courses that are valuable but rather that environment of being with other people that makes you a better person, gets you thinking outside the box, and prepares you better for the world.

I graduated from Rensselaer Polytechnic Institute in New York in 2001 with a BSc in computer science. I started working for GE in New York for about two years as an application developer, then moved to a start-up in San Diego, California, as an application developer but I also worked in quality assurance, product development, project management, program management, sales engineering, support – everything you can learn from a start-up environment. The company – which was called Telephony@Work – was acquired by Oracle around 2006 and so I joined Oracle as a program manager, first integrating the start-up's product line into Oracle and then moved to the Siebel CRM product line, where I was the program manager for the product releases.

I spent about five years at Oracle but I always had a craving for going back to a start-up environment as well as to a technical environment. I continued training myself on the side when mobile applications were booming for iOS and Android, so that was my hobby. At some point, I said corporate life was enough and I quit.

I am originally from Turkey and I moved back there after 15 years in the US to start a digital company doing mobile loyalty, mobile payments and everything that you do at a store through your mobile phone. I did that for about three years, working with the third-largest retailer in Turkey and winning two awards in London for the product that we built in 2013 and 2014. Then my big exposure to machine learning started.

I got an offer from the UK government for the Global Entrepreneur Programme (GEP), to move to the UK and continue my career as an entrepreneur there. I accepted that in 2015 and moved to London, where I started a company called LeftBrain Technology. It is a marketing automation platform for ecommerce business, which focusses on 100% automation of customer communication for any online business. The company has performed well, increasing sales from anywhere from 20% to 80% for its customers. It is a business that is on the side, so I am not fully immersed in it, I have people managing it. About a year ago, I wanted to get into more hardcore artificial intelligence, so we co-founded DeepZen, which has been 90% of my focus for the past year. It uses the latest artificial intelligence technologies to imitate the human voice, with expressivity and emotion.

KEY TO SUCCESS

Curiosity, self-teaching, collaborative cultures, and applying your knowledge.

I was always good at statistics and passionate about it when I was younger. I always loved to see statistics and data of everything – football scores, anything – to the point where my friends would make fun of me. My formal academic training was on the technology side, but I have taught myself data science techniques such as coding and programming, and mathematical and statistical modelling too. Just being good at maths and statistics was not going to give me the ability to do what I wanted to do. I had to self-teach new languages so that I could apply these statistics and methodologies in the code itself to build an end product.

I took online courses. I explored our system's back end, which was a good way and a simple way to start with machine learning. I bought some books and went chapter by chapter and tried to teach it to myself. The good thing is that we are in the age of the internet and there are so many open source projects like GitHub and other things where you can go in and try things out. I push people to read white papers, as well as do something and get their hands dirty. That is the best and quickest way of learning.

The PhDs that I am working with are very smart, even smarter than me, but they are good at applying methodologies, not necessarily coming up with new ones. I try to come up with new methodologies and then my team applies them. The product knowledge I have, and the experience of building a product changes everything because I think from the end product perspective. Whereas somebody who has been going to school for 25 years, who is very smart, usually thinks in a narrower fashion because he/she mostly focused on theory throughout their careers. That is the key differentiator. You need to have experience of applying your knowledge from the PhD in building an end product in real life, in order to be able to think outside the box.

One of the things that I love about being in the UK is that collaboration is very important. It is a more vicious situation in Turkey and in the US, especially in Silicon Valley. For lack of a better term, Silicon Valley is all about stepping on each other so that you can move to the top. The collaborations might not always be honest, from my experience. In the UK, it is a smaller environment with a lot of smart people. The top universities are very good on artificial intelligence and I have the feeling that people like to help each other more here than in the other places that I have been to.

WHAT ARE THE NEEDS?

Cross-sector networking and knowledge exchange, international and multi-sector experience, learning on the job.

It is critically important for start-ups to work with people in academia. We would love to hire five or six people as an intern or similar, who are going to school on a part-time basis, because it is fresh information: they are learning the latest things. I wish that was easier to do. I think there needs to be more specific social activities or conferences that bring academia and start-ups together. Start-ups are obviously limited on budget, so we would love to get academics involved – not postgraduates but people who are at an earlier stage – and we might not be able to afford to. It would be great if the government could create funds somehow to take some of that cost off the start-ups' back, so that costs can be shared between the start-up and the government, to expose academia to start-ups. Everything that you hear that is leading right now started in the past 10 years or so as a start-up. That is where the future is, because start-ups can move faster than corporations, there is more talent and younger people, people with more energy.

DeepZen today is a team of 11 people across four or five countries. We have people in Brazil, London, Oxford, Turkey and in Tunisia. We have somebody joining from Romania. I encourage people to move around and work in different places and to work remotely. I think it makes you a better person and a better researcher when you are exposed to more than just sitting in a classroom. In a specific country, when talking about statistical models, it is important to be exposed to different types of data in different industries and countries, even in different cities, because when the data is dependent on the consumer, it can vary a lot, and that means different statistical models might need to be used. I also recommend that people move as much as they can, to different companies or industries, for faster development.

In our case at DeepZen, because we are a start-up and in a competitive space, we need to move fast and we are very busy. I try to hire people who are self-learners and who are willing to put in the extra hours to teach themselves, because we do not always have the time to handhold. I do not care if they are not very experienced, but if they have the ability to learn fast and adapt without too much handholding, that is good enough for me. I do not care if the person has a PhD or has not even graduated, as long as they have the mindset and the capability to learn.



CASE STUDY

Revolutionising the NHS through data science

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

CAREER HISTORY

Medicine to academia and data science

I went to medical school in Edinburgh. It was a five-year course but I did an extra intercalated year in Pharmacology, and from that point on I was interested in statistics. I had the opportunity to go to Columbia University to work in the biomedical informatics department. There, I started to get switched on to the idea of data science applied to medicine electronic records. We had a software developer who wrote a Python script to go through the electronic records in New York Presbyterian Hospital. But I was struck by the fact that we could not really talk to each other. It was difficult for him to convey the statistics that are commonly used in medicine.

Throughout the end of medical school, I taught myself Python using Codecademy, which is a highly accessible online resource, with a view to starting an evening course at General Assembly. They plug skills gaps in the market and courses are intense courses to get professionals quickly up to scratch. Their Data Science course worked best for me. It was one of the best decisions I ever made!

Soon I moved to London to further advance my data skills, alongside my career in medicine. At this point I was working as a junior doctor at UCLH and studying data science in person, two evenings a week. Luckily the tutor was from biomedicine so a lot of his examples were using open source brain tumour data. I could start to see that this could be really useful for medicine. But I noticed the gender imbalance. It was a class of about 30 and there were two women, including myself.

Now I work as a Senior Research Associate at UCL Institute of Neurology. I am currently working on predicting whether patients will attend their MRI scan appointments. I do this in the same way a mortgage adviser would look at somebody they would potentially give a loan to and work out a risk assessment for them.

KEY TO SUCCESS

Data access and open source models.

I came in to this from the point of wanting to improve the NHS. That comes from just being a doctor and feeling a bit frustrated. There is something to be said for having satisfaction that you are doing something good for the benefit of society. Non-attendance costs the NHS £1 billion every year. I helped build a model that can predict which patients will not show up for their MRI scans, which is currently the best performing in the world. If we can get machine learning models like these implemented, I think we will see a surprisingly early return on the investment of data science training.

We also have amazing data in the NHS and that is definitely a resource that is worth staying for. In addition, a lot of the models that are very useful for this are open source and accessible to someone who is relatively new to data science.

Having said that, I do see why people go to industry and it would be totally understandable if people were attracted by the increased pay.

Build foundational skills and bust the ‘maths-genius’ myth

A prerequisite for data science is having some sort of computer literacy. You can apply lots of different models to different types of data, and you can only get to the higher dimensional more complex models if you are able to code. I think that currently is a pre-requisite, but I do not think that it will be in the future. I am sure there will be platforms available – you will be able to apply data science to your data without having to code from scratch.

It is not that common for my generation even to be able to program. Although there are loads of really great online resources, I think it can be difficult to see why it would be useful. If I were to have my own medical school, I would definitely have a programming section! You do not have to teach everyone, but you can just have it as an extracurricular option.

Another thing is getting people to have the confidence to think that they can do it. This is seen as not something for medics, or for people who are not from a maths or computer science background. There is this common misconception that it is really difficult when it is not. It is more language-based than mathematics based. There are loads of really great resources such as Code First: Girls.

I would say after learning to program, the next stage is learning data science! You can speak the language but if you have got nothing to say you do not get anywhere. Again, people think that this stage requires very advanced mathematical knowledge, like matrix calculus and you have to be a university level mathematician or physicist, but that is not the case! You can build models almost intuitively at first, and you can just see what works. The open source data science community is so good. It is amazing that all of this human effort has gone in to creating these open source packages, and then people just put it online for free.

I am not sure if I could go back to medicine now. It is not really acceptable to take long career breaks, which comes primarily from a safety point of view. I think it would be nice to have more accepted routes to come back in like retraining programmes. Perhaps these exist, but they seem a bit adhoc and I am not quite sure how to access them.

The point at which you want to start customising your models, the point at which you think ‘I am not quite sure I have built this model optimally and I want to try and do it better’, that is the point where you need to get the mathematics skills. That is tricky; that is where I am trying to skill up at the moment. Luckily I have got a team who are really keen to help and suggest homework, but there are also loads of great online resources, like Khan Academy. I think if you do all the earlier steps you naturally come to the point where you want to know more about the mathematics. I do not think it requires any special dedication or geekiness. I think it is just, ‘I have built something and I want to make it better’.

CASE STUDY

Leading the data science profession within UK government

Nick Manton is 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.

CAREER HISTORY

From physical sciences in academia, to leading data science in government

I graduated from my PhD in materials science and worked as a postdoc fellow at Stanford Research Institute for a year. When that finished I went into operational research. It seemed a good fit in terms of the transferable skills from my education – the analytical side, the problem solving, the statistical modelling – so I joined the Government Operational Research Service in 1999 and worked my way up as a government analyst.

A sizeable proportion of data scientists and operational researchers within government have postgraduate degrees in physical sciences, such as particle physics. It gives you a good grounding in statistics and mathematics, which are the kinds of skills that you need to do statistical modelling.

I worked on various different analytical projects and statistical models in the Department for Social Security then, after time at the Home Office working on performance management systems, I segued into a digital role, building digital performance management systems for the police and probation services. I joined the Home Office Data Analytics Capability and my job was to lead the lab and recruit civil servant data scientists to build up capability.

After a couple of years, I was asked to head up a digital and data team in the Cabinet Office. I led that team for 15 months, working directly to No 10 on the race disparity audit, which was one of Prime Minister Theresa May's first policy initiatives. Then I moved on to my current role as Head of Data Science within GDS where I run the data science function with a team of 9 people and co-lead (with ONS) the data science community across government.

There is a real intellectual curiosity around using data and the tools, techniques, algorithms of how we can exploit data more effectively. It is only in recent years that those tools and techniques have become readily available and accepted for use by government.

One of the things I like about the data science domain versus analysis is that there is a research focus in data science. Within the Cabinet Office, most of my team work on GOV.UK looking at the user journeys and we use network science to improve user journeys, which is a common problem across any industry that has a significant web presence. Across government, there are some massive challenges and some really interesting work predicting recidivism within the prison and probation system or criminal justice. All sorts of really interesting problems that you would not necessarily see within, say, a fintech company.

Particularly on the digital side, we tend to be more about building products than necessarily doing one-off analyses – so producing a number or a value or a chart for a minister or a customer. User research is incredibly important in that. If we are building a product, we might look at software development, software engineering, security, product management, delivery management, to keep those teams running and energised.

Moving into the civil service, that variety of backgrounds and interests, but also the interaction with policy and operations, meant that it was more diverse than in academia. Generally, I have always had that sense of being valued as an analyst within government, no matter what the project. Moving to government, particularly in a semi consultancy type role, meant that the results were a bit more evident. For example, the differences from a string of numbers about pressures and gasification metrics to the number of people getting a Social Fund benefit.

Access to data is also a problem within government. It is getting better but still there is a much lower risk appetite for experimenting and trying new things with data. That then tends to slow the innovation and R&D that you can do in government, so having people who can go from government into private sector/academia can certainly benefit.

We do run a data science accelerator scheme within government and GDS is one of the big hubs. Analysts who have a business problem and want to learn some data science techniques can apply to the scheme. If they get in, they are given a mentor, an experienced data scientist from the community and access to a decent laptop unconstrained by corporate IT restrictions. They also need to have their head of profession and manager sign up to say that they can take the time off each week.

Enabling movement

Working in the private sector is going to enhance knowledge. The private sector, typically, is a bit more advanced in terms of a lot of the tools and techniques than in Government, which is slightly lagging behind. Pay is the obvious barrier to doing that. What government can offer is access to tackling some really interesting problems.

I would like to see a mechanism that would allow any data scientists to do a shadowing exercise or a secondment (three month, six month or 12 month secondment) at the Alan Turing Institute or the Ada Lovelace Institute. The challenge is backfilling that person, but if I give you a data scientist, you could get a research fellow from ATI to come and work in government for six months. Our data scientists would benefit from being taken away from the pressures of delivering to a customer and having a bit of latitude to do research. Mentoring is also good and it helps the mentors as well, because acting as a mentor means that you need to think about problems in different ways.

Accessing data engineering skills

As technology generally improves, the opportunities to collect more and more data are increasing. Good data scientists are inquisitive about the data and will be active in thinking about where it comes from, how to get it and how to transform it. Unfortunately big industry is paying significant premiums for this skill and experience, and making it hard for government to compete in the market.

But perhaps the bigger problem is getting people with good data engineering skills. People who have knowledge and experience of handling ‘big data’ – huge, unstructured datasets, data sources and/or very rapid streamed event data. These are massively high volume, high velocity datasets that may need to be stored, processed and analysed in real time. The plumbing and the building of the pipelines to handle them is a crucial and critical skill for government. The new data coming along could be social media data, or sensor data, basically live data in real time from borders, and then using Hadoop or Spark (or any of that family of big data technologies). We, data scientists, will have to do the data engineering role because there is no one else to do it, which is then slightly suboptimal for what we should be doing.



CASE STUDY

Using data from the web to understand human behaviour

Chanuki Illushka Seresinhe is the Lead Data Scientist at Popsa, a start-up using AI to automatically curate photo content into designed physical products; and a Visiting Researcher at the Alan Turing Institute.

CAREER HISTORY

Design and digital media to behavioural economics and data science

I initially became interested in the internet from the website point of view. At the time, in the late 1990s, most websites were basically centred content on one page, and this was still really exciting. I studied digital media at the University of California, Santa Cruz. Because of the novelty of the web at that time, there were no courses you could take to learn how to design on the web, so I kind of evolved as the web evolved. I ran my own digital design consultancy, but at a certain point the web became quite formulaic. Everyone had figured out websites and there was not anything especially creatively challenging or fun to do, and I was trying to look for something new that was intellectually challenging.

I decided to return to University, and while I did a Masters in behavioural economic science at the University of Warwick, I saw this really interesting opportunity at the Warwick Business School, where there was the possibility of using data from the web to understand human behaviour. Suddenly there was this light bulb moment, because I finally found a new angle on what to do with all the data we all were producing on the web. We would not have data science as we know it today if it were not for the all the data generated on the web through our online interactions. It really inspired me, and that is how I got into data science. I went on to complete my PhD in data science at Warwick, before taking up a post as Senior Data Scientist at Channel 4, and then moving to become the Lead Data Scientist at Popsa – using AI to automatically curate photo content into designed physical products.

Logic, empathy, communication, and project management

You have to be very good at hypothesis testing, for example, and to be very logical with how you 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. If you do not know how to communicate your results, the fact you come up with a very interesting study might not really make any difference at all.

You need empathy as well. You are often working with data that has a lot to do with people's lives. You have to be really careful about how you interpret that data and how you understand it. Some of the conclusions you come to are not necessarily black or white. You have to have an additional skill set to try to understand the data and how to make sense out of it. We have to be sensitive as to what we are uncovering. We need to be careful to question our models and make sure we are actually reaching the right conclusions.

I have gained a very good reputation for my research, and in that sense I feel valued in academia. But in my day to day life I feel more valued in the commercial world. I think this is because I have a very broad skill set, and it feels like the commercial world accepts that a lot more than the academic world. Academia looks at a couple of skills but they do not necessarily look at everything that you bring to the table, which is a shame, because to be a good data scientist you really do need a hell of a lot of skills. When I was doing my PhD, the skills you get taught are basically about how to do research. Traditionally, in academia they do not value, for example, how good a communicator you are; they do not necessarily look at how good you are as a project manager.

However, if you have a lot of experience project managing, then you are more likely to get your research done faster. I have achieved a lot of press for my research, and a lot of that is because I know how to distil my findings into ideas that are easily understandable and interesting to a public audience, because I had another career before coming to academia.

When I decided not to do academia full time, financial reasons were an important consideration. Another big issue I have is that academia tends to be quite traditional. I had an entire career before I went back to university, so I am really not in a position where I could start from the bottom up again. It doesn't seem easy to move sideways in academia, whereas in the commercial space it is a lot easier, because people take into account my background and other experiences, and have let me move up a lot faster than I felt I could in academia.

Data science is quite new to the commercial world. A lot of it is coming from very good software developers working in data science, but sometimes it feels like they could benefit a lot more from collaborating with universities to get that scientific part up and get to grips with how they might be using data science – particularly things like deep learning and artificial intelligence. In that sense, I think that universities know how to do that a bit better.

In the commercial world you do not have that much time to research. You need to come up with answers a lot faster. You have to prove business value, and it makes you quicker at doing your data science, and working with more efficient methodologies, because you do not have six months to explore a topic, as you might in academia. It forces you to think on your feet a little bit more and to try to hone in on the application of your research, so you are not just doing data science for the sake of it. But you can imagine how such structure might be valuable for academic research: for example, you might want to provide value for society by coming up with very solid conclusions.

Ethics training, diversity in data science, grant funding for collaboration, and coding training in academia

It is really important that data scientists are trained in data ethics. But it is also crucial to have diversity in data science: we also need to focus on how to upskill, retain and share data science talent in women and underrepresented minorities. In data science it can feel like there is a “bro” culture. It all is very subtle and I don’t think most people these days are overtly trying to exclude anyone. But I wonder how much of a part this plays in the dearth of female data scientists we still have today, so I do think it is something that needs to be addressed. I think it is particularly relevant to data science as we are creating algorithms that impact many different types of people. Having a diverse set of data scientists means that inadvertent biases that might creep into algorithms might be thwarted, as a diverse team might be more likely to pick up or be concerned by those issues.

Once you have left academia and you are not publishing high quality papers, it is very difficult to go back to academia. On the other hand, in the commercial world, they are not very keen on you doing independent research. If there were more funding for projects that required one commercial partner and one academic partner, that would be interesting. By letting the researchers come in as consultants through something like grant funding rather than consultancy fees, it gives them the exposure and experience of working in industry without having to leave their academic careers, but it also allows industry to see the value of how research is conducted in academia.

One of the challenges of data science in the commercial world is that you have to be a very good programmer. That is not something I was necessarily trained in in academia, and I do wish that universities paid more attention to that. The level of code required in the commercial world is a lot more serious than you might need for academic research. When you are doing your academic work, you tend to work by yourself and then it really does not matter that you have messy code. The problem then happens when you start to collaborate with other people. Then you have to write code in a way that somebody else can understand. I think it is important to teach that skill, because we want to have reproducible research as well – so that anybody can understand what you have done and replicate it. If I had been taught to write code like that at university then it would have allowed me to have more collaborators, and to share my code for more critical reflection. Collaboration can also introduce that extra scrutiny.

CASE STUDY



Contributing to academic research through the creation of software

Dr James Hetherington is Director of Research Engineering at the Alan Turing Institute. He leads the Turing's staff Research Data Scientists and Research Software Engineers to help ensure that their algorithms and analyses meet the highest standards of software architecture, reproducibility and quality assurance.

CAREER HISTORY

Academia to start-ups and back again

My PhD was in Cambridge in theoretical particle physics and then I moved to the life sciences for postdoctoral work. I completed a five year postdoctoral fellowship at UCL on a mathematical and computational biology project. BBSRC were aware there was a lot of mathematical, quantitative, computational work emerging and so they were funding people to move into this area. Unfortunately, I wrote too much code and not enough papers during that postdoc, so there was no academic position for me at that time!

After leaving UCL I worked for MathWorks, an American corporation that specialises in mathematical software, where I learned the craft of software engineering. They had a strong internal training programme and they took people who could understand the mathematics. That balance of maths and stats with software engineering is data science for me.

Then I went to a London based start up which modelled human environmental impact called the Avoiding Mass Extinctions Engine. We made an API and people could post real-world questions to it, e.g. "how much pollution does my aluminium factory make if I make x amount of aluminium a year?"

In 2011, I came back to academia and joined UCL doing mathematical and computational modelling of biological systems, this time on brain blood flow and aneurysms. I had not been publishing while I had been out of academia. But my group leader, Professor Peter Coveney, realised that he needed people who were experienced software engineers for his broad area of interdisciplinary research where high performance computing plays a major role.

Championing Research Software Engineers

Lots of the software that emerges from research projects is not usable by anyone other than the PhD student who wrote it. Working with the Software Sustainability Institute, I and others led the creation of the UK Research Software Engineering Association. We have an annual research software engineering conference, which is in Birmingham this year, with 300 attendees.

I also worked with senior management at UCL to create a Research Software Engineering Group, which I led between 2012 and 2018. I am proud to say that it was the first group of its kind in a UK university, and grew from just me to a team of 12 in that time. The group collaborates on research grants in order to create really high-quality, well-engineered software out of research activities.

It is really important to create a space in academia for people whose contribution to research is through the creation of software, rather than papers. If your contribution to research is primarily through the creation of computer software, how do you garner academic career credit?

Recognising diverse outputs and building teams

The research engineering group model has since been taken up by many other research intensive universities in the UK including Manchester, Bristol and Southampton. One thing that makes the groups successful is that we take both service and scholarship seriously. This model should be rolled out further and it will also be interesting to see how alternative outputs are included in the REF.

The last thing I would say is that team-based research is really, really important. There are implications for how we prepare people to be part of those teams. I want to champion existing mechanisms such as the STFC data science doctoral training centre in UCL and of course, the UK Research Software Engineers Association.

CASE STUDY

From Silicon Valley, US to Silicon Fen, UK

Dr Damon Wischik is a Lecturer in the University of Cambridge's Computer Laboratory and former recipient of a University Research Fellowship from the Royal Society. From 2011–2016, he worked as the chief data scientist at Urban Engines, a Silicon Valley startup now part of Google.

CAREER HISTORY

Career history: academia to industry and back

My PhD supervisor, Frank Kelly FRS, is renowned for his work on mathematical modelling of networks. At the time he had made a major contribution to the theory of Internet congestion control, and that was the topic of my early research in the maths department at Cambridge. In 2005, I was awarded a Royal Society university research fellowship, and I joined the networks group in the computer science department in UCL: I wanted to work out why all this brilliant theory wasn't actually being used. We turned the theory into deployable code, and it's now in widespread use – including in Apple's Siri.

In 2011 a colleague in Stanford was starting a project on congestion control, this time for urban traffic rather than Internet traffic. The project was exciting, and I wanted to see what Silicon Valley was like, so I jumped at the chance to join his startup. I moved to Palo Alto, and ended up leading the data science team of ten people. We worked on behavioural nudges, and on understanding how users interact with a city's transport network.

I was with the startup for five and a half years. We are living through a crucial time of how technology is shaping the world, and being part of a Silicon Valley startup was a fascinating way to experience it. (To cap it all, I was at home one evening watching the HBO sitcom "Silicon Valley", when it showed a shot my apartment building!)

In 2017 I came back to academia to take up a position as a lecturer in data science at Cambridge. I continue to work on networks, behaviour models, and congestion, but now (I hope) with a much clearer picture of what's worth working on.

From the point of view of academic career development, my six years in Silicon Valley were a waste: no publications, no pipeline of academic research, nothing tangible on the academic career ladder. But it was vital experience, and I am grateful that the Computer Laboratory at Cambridge recognised it. This kind of experience is needed in academia.

Identify problems with impact

I moved to Silicon Valley because I cared about the impact of my work. In a startup a key metric is the number of “monthly active users”, and it was thrilling that I had 300,000 users interacting directly with my software, my ideas. A startup lives or dies by whether it’s identified a real market, and whether it’s created something gripping enough that people want to use it. Academics are always being asked to write impact statements – but the link is usually attenuated, not the thrill of direct impact nor the struggle to make your work compelling.

A perk of startup life is that you can spend your time thinking about the ideas, whereas in university life there are endless lectures, committees, and grant proposals. On the flip side, there are some big ideas that grow slowly, from flashes of inspiration that come to you on the bus or the bath (as my supervisor used to say) – and I was drawn back to academia because I wanted space to develop these ideas.

Data science apprenticeships

My startup, like so many others, struggled to hire data scientists. At Cambridge I am involved in changes to the curriculum, at all levels from first year undergraduates to masters students, making sure we teach the skills that students need for data science and machine learning. So far we have concentrated on computer science undergraduates, but we’re beginning to plan how data science skills might spread through the rest of the university, and how a computer science department might help.

For serious data scientists, I think the most promising route will be some sort of “apprenticeship masters degree”, one year of taught masters course plus one year of practical work. It should be like an artist’s studio, where apprentices have enough time to become invested in a project, but not so long that the project is the be-all and end-all. Data science is more like a highly skilled trade or profession than it is an academic discipline. A PhD isn’t the best route, since it encourages a single-mindedness that is unhelpful in data science. It’s also not good to go straight from an undergraduate or masters degree into industry, since it’s too easy to be diverted away from data science and into software engineer. Some companies are trying to set up this sort of apprenticeship training – at Uber and Microsoft they call it a “data residency”. Universities have a duty too, particularly to train the data scientists who will go on to make the world better, not just those who will improve some company’s ad revenue.



CASE STUDY

An analytics apprenticeship at the UK's largest independent producer of official statistics

Alexis Fernquest is a Data Scientist at the Office for National Statistics (ONS), and former Data Analytics Apprentice.

CAREER HISTORY

Financial sector to the public sector

I started with a degree in economics and then went on to a graduate scheme in a small private company that was selling technology. I was there for a couple of months before I joined a larger organisation, Lloyds Banking Group. I started in household claims and I worked on the telephones for a little while, dealing with queries. I was very quickly drawn to the analytical side so I started getting involved in some resource analytics. I flourished in that area and really enjoyed it.

Then an opportunity came up in the Office for National Statistics to get involved in a data analytics apprenticeship. It was the first of its kind in the UK. They were promising a huge amount of training involving different languages and coding. It was exactly what I was looking for. The apprenticeship officially lasts two years but I completed it ahead of that. I was made permanent and then promoted to data scientist.

I have had the benefit of seeing both the private and the public side, so it has been quite an interesting experience. When I was working in private sector I did not feel that they were invested in giving me extra skills. I have always wanted to learn to code, but because they did not necessarily use coding at the time, they were not that interested in the investment. At the ONS they had this programme where you could move into academia again and learn, and they were willing to train you up to get you into that role.

Teams and communities

When I worked in Lloyds Banking Group, they were into agile project management which makes you work well across teams and have oversight of what every department did.

When I joined the public sector, it was sometimes quite difficult to get your head around exactly who did what, but I also felt a lot happier to get involved. I feel confident to raise an opinion on how we could be doing things differently. At the ONS they really encourage you to move around every two years to different departments so you can work on all different types of analysis. It could be economic statistics, where I am at the moment, or methodology. They try to vary it so people do not get siloed.

As a university student, I did not really have to do much team collaboration, whereas in the apprenticeship I definitely had to do quite a lot. There was a whole unit dedicated to how you work with other people in an office. I think it was really good, because we had a couple of projects that we had to do amongst other apprentices, and then we had to go out to other departments and get to know them. It was a really good way to get involved and work in a team.

Most employers are too traditional in their approach to hiring people. I think if you take people in with an apprenticeship, you pay for them to go through that training and they are going to be loyal to your company and the retention rate is going to be higher. Secondly, if they are going to work with people across different sectors, you are going to build a really good network. When you graduate you will have a cohort of alumni apprentices who could help others. A first year apprentice could be stuck on a problem, and use Slack to seek an alumnus to help them out. I think community meetups are important so you can talk to people from other areas. At the moment there is not a channel in place where you can easily get in touch with people working in your profession.

Build interdisciplinary skills

Data science is an interdisciplinary career. You have to have maths knowledge, coding and computer science, but also be curious and want to hack things. I work in a lot of new and upcoming technologies, like Spark, dealing with parallel processing. I see the architectural, techy side of things, and I would definitely say I am more of a data scientist than a mathematician. Coming from university and going to work in Lloyds, I had to develop my communication skills. For the apprenticeship, I really had to increase my statistical knowledge.

In 10 years' time you are going to have all these people coming out who are already excellent at coding so you have to be quite resilient and open to change. There could be a lot more done to develop flexibility, resilience and resourcefulness. The whole point of data science is that you are not amazing at everything, but you have a little bit of knowledge of all these different things and you can draw it all together.

Working at the ONS I feel proud about the work I do and I can actually see how it impacts policy, which is really exciting. I am starting another degree in September in data science, one of the first of its kind in Wales.



CASE STUDY

What will transport within cities look like in twenty years? Understanding government data science challenges

Frank Kelly CBE FRS is Emeritus Professor of the Mathematics of Systems at the University of Cambridge. He was awarded a CBE for services to mathematical sciences in 2013 and he is a Fellow of the Royal Society and the Chair of the Society's Advisory Committee on Mathematics Education.

CAREER HISTORY

Academia to public sector and back

After my first degree in Maths at Durham University, I worked for a computer/management consultancy for 18 months, before starting my PhD in the Statistical Laboratory at Cambridge. This short interlude in industry strongly influenced my choice of research topic; mathematical models of large systems; with computer networks as an early motivation, and gave me a taste for interactions with industry. I have pursued a career in academia, but with several periods in industry or government, and with joint papers with colleagues in Bell Labs, BT Labs, Microsoft Research, Huawei and other companies.

My main research is in random processes, networks and optimisation, with a specialist focus on applications to the design and control of networks and to the understanding of self-regulation in large-scale systems.

From 2003-06, I was the Chief Scientific Adviser at the UK's Department for Transport. This was a fascinating and rewarding insight into Government decision-making. Climate change, pollution, congestion charging, airport planning, railway signalling were just some of the areas where science in the broadest sense, including technology and mathematical modelling, needed a voice at the highest levels.

WHAT ARE THE NEEDS?

Finding ways for universities and government to collaborate

Looking to the future, many of the most challenging applications of data science arise in areas such as transport, health and education – where Government has a key role to play.

What will transport within cities look like in twenty years? Most of the physical infrastructure that will be present is already there as it takes a long time to change. But how will it be used in the future?

We should expect the decreasing cost of car travel, driven by the economics of electric and autonomous vehicles, to require a radical rethink of how scarce road space in cities is allocated. This will be enabled by close integration of traffic control (traditionally run by cities) with route choice (traditionally user centred). This is a major data science and AI challenge for government, innovative companies and universities.

In my experience, the boundary between universities and government is potentially difficult as 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.



CASE STUDY

Applying data science to end extreme poverty

Kevin Koene was previously a Junior Data Scientist at The One Campaign, a campaigning and advocacy organisation taking action to end extreme poverty and preventable disease, particularly in Africa.

CAREER HISTORY

From finance to data science in an NGO

In terms of my data science career it is still in the very early stages. I did a Bachelor's in economics and a Master's in finance, which is heavily reliant on data, both in Rotterdam. I have always been interested in shifting towards data science and London offers more roles in international relations and development.

The opportunity arose when I saw the vacancy at The One Campaign, when they were hiring a junior data scientist. It is not the classic path to work at an NGO in terms of data science, but that is where my passion has always been and I wanted to find a way of using that passion in my work. Data science combines statistics, mathematics and computer science, it is a fusion of the three.

We are focused on having a data driven culture within the global policy team. We analyse data on poverty, focusing on areas where it is the most extreme, allowing practical interventions to be followed up. This can be challenging as the data in the development field is not always very detailed. There are a lot of data gaps, especially for historical data, but this really motivates me and my team to find ways of using the data and improving the data that is being collected in the future. There needs to be sound estimations and regressions that we run with the data. Sometimes I am called into projects and when I look at the methodology I shake my head because it isn't as robust as it could have been in, say, academia.

In the Abuja office we have contact with the government in Kenya, they give us a dataset on gender based violence and then we find the facts and the background of this data. There are also a couple of long term projects that we have, like working with the OECD's official development assistance data, which happens twice a year, so we are writing scripts that automatically pull the data, shape it and prep it for our needs. We are also pushing for open data, most of the data that we work with is publicly available.

Informal learning and mentoring, backed with experience

I do not have a background in international relations, whereas most of my colleagues do. On the other hand, my experience in finance was helpful and it was good to have people from different backgrounds. Our data scientist and my mentor, Kate Vang, is also from an investment background. She mentors me and gives me support in making the change from the finance world.

My academic experience was quite competitive because it was in finance. Now, going into the NGO, international development field, it is a lot more collaborative. In some senses, it has been a 180 degree turnaround from the very private orientated finance world to a public sector NGO. In another way, it has just fuelled my passion for data and data science.

My use of STATA during my thesis life is also benefiting me now, but I was never aware of the possibilities of R and Python. I gained that experience by stepping outside academia. Things like STATA need licences and accounts. All those still will have a monopoly in universities. That should change. R is open source and it is free.

What I really like about data science is that the community tends to be very open and helpful, so I take online courses on edX or Udemy while working and I am able to get help online if I am stuck at a certain point. On GitHub there are a lot of public code repositories, which serve as a sort of library for code and files and allows me get inspiration from other people's work, and I joined online communities for R and Python. GitHub is also good for recruitment because can see what you are working on. A lot of people put anything that is not proprietary on there to share their work and build a portfolio.

However, I think the most important part is having proven skills. In an interview, they will have a feeling about whether you are familiar with use cases, or whether you just did an online course that teaches you the basics. That makes a crucial difference between only having done an online course, rather than having applied it to the job and being able to adapt to circumstances.

You will never just be asked, 'Have you used R before?' or in finance, 'Have you evaluated bond prices before?' There is always going to be a next question, 'And how did you...?' That is why, especially in data science, it is good to have something practical, whether it is code or a project or something that you can show. That is where, I guess, GitHub comes in, where a lot of your work is public. I know, that a lot of people put anything that is not proprietary on there to share their work.

Embed data science across disciplines

Data science is helpful to almost any science. I know that I would have taken a data science project on the side if I could have. The important point is not having only one module, but having it influence your other work. Maybe have a module in terms of preparing for the thesis and then apply those skills to your thesis. I had a module called 'Qualitative methods in finance', in which we used MATLAB, so it was applying MATLAB to finance, but it was a standalone module and thereafter I have not used it, plus it's not the most user friendly software.

There is an increasing number of bachelor and Master's degrees in data science, but other degrees should incorporate data science more. Machine learning is going to be a big part of the future. Skills are in short supply now and there is only going to be increasing demand. Data mining too, because we are gathering so much data and we need to improve the ways we are working with it.

CASE STUDY

The skills that define a data scientist

Milton Luaces is a Senior Manager at Accenture – Applied Intelligence.

CAREER HISTORY

Spanning academia and senior management across sectors

I am a computer science engineer with an MSc in statistics from the University of Sheffield and a PhD in artificial intelligence at Universidad Politécnica de Madrid. I have more than 15 years researching, publishing papers and working in the industry, exclusively in the data science field. I was already working in the private sector while I did this postgraduate work – the data I used for my projects came from industry.

In a previous life, I was a lecturer and researcher in university. My reason for doing academic work was for the innovation and building practical, real models with real industry business problems. I see myself mostly as an engineer, so instead of just doing research and publishing papers, I want to build things and do large deployments in big clients. That is really motivating for me and why I decided to move from academia to industry.

I am in financial services right now, but I have also had a lot of clients in health and public services, and supply chain. In industry, what they are interested in is the value drivers: “what can you do under constraints?” It is not science anymore; it is a completely different way of thinking.

I feel more valued in industry than I did in academia, maybe because my approach fits the requirements and the expectations better. Expertise is very valuable – especially for large companies and because of the value you add you immediately go to the top of the decision makers. Additionally, compensation and benefits are higher. In academia someone could spend years doing fantastic research and could never be recognised – or the importance of the work is only recognised after they have died. But companies have very few people in senior data science positions. And so in industry you have instant recognition if you add value to your client.

A portfolio of technical skills, business skills, and soft skills

My view is that there are four skills that define a real data scientist: statistics; machine learning; computer science – not just the ability to write snippets in a language but software; and then big data, because data in the 21st century is big data. What is in particularly short supply right now is having the four main skills together.

Then you could also need industry acumen, but the industry acumen does not need to be at expert level. You just need to understand how the business works. If you are an end to end professional in data science – such as starting from the data, understanding, analysing, modelling, validation and application – that means you have to discuss with the client their business problems and then go back to your team and brainstorm with them how to solve this problem until it deploys and is working.

This a real challenge when you jump from one industry to another. But once you specialise in a particular industry sector, there is not so much variety – you start to see that the problems replicate themselves more or less. There are data careers where the industry acumen is quite important, for instance in pharmaceutical results, or data engineering.

One of the important differences between industry sectors is the culture. There are sectors like IT companies, for instance Silicon Valley based companies, which are open minded to innovation with strong funding for research and development. On the other hand, there are more conservative sectors, for instance financial services, where regulations play a bigger role in defining how clients can innovate.

Project management in data science scenarios is completely different to traditional project management. The iterative process is riskier; you have uncertainty all the time and may have to run different things in parallel to see what works. You are trying to innovate so you can't rely on previous project management templates. Some people want to work in data science because they think there is no politics in it; they are wrong. Negotiation and consensus building are necessary soft skills. You have to sell your project goals both internally and to you clients, because sometimes they are a leap of faith that there is value in the data. You also need to make sure you are connected to the teams that will deploy what you prototype. If they do not trust in you or they do not have aligned interests, it will fail at the end.

Career breaks and knowledge exchange for senior scientists, risk appetite for innovation, and data access

If an academic goes to the industry for some time, when they come back to academia they cannot have the same level that they had in the past, and that is a problem. If you can have a leave of absence from academia, keeping your career level and spend maybe one year in the industry and come back, you could enrich your own research with knowledge in the field and share your knowledge with others. There are many companies that allow their staff to have a leave of absence. Your salary would be suspended during your sabbatical but when you return you keep the same role and salary that you had when you left. Another good way for sharing knowledge across academia and industry is having panel or roundtable events on specific hot topics from time to time for data science leaders at senior levels to discuss, share and take advantage of community knowledge.

I have lived and worked in four countries. Data access is a real challenge right now but it varies across world regions. I think there should be a balance between the privacy rights of citizens and the need to take advantage of massive datasets, especially in governments. I do not think we have clearly defined the right balance between regulation and innovation on data yet.

In the UK the level of data science is really high compared to the European context. The English language could be an advantage, because of the close relationship with the US. The City of London, the financial centre of the European Union, is a big attractor for talent. All of this adds to the leadership the UK has within Europe in this field.

On the other hand, in some ways the UK is a step behind some US environments for data science, especially the West Coast. The level of academia is quite similar for the top universities in the UK and, for instance, in California. But the UK is more intellectually conservative than the US. For innovation, you need to go ahead with your ideas. Some of them could be reversible if they are wrong, but you need to take risks and embed this really well to do the best innovation. It is perfectly possible in the UK to reach the same level as the US in delivering data science innovation, but there is a culture of risk aversion. It is in the DNA of Silicon Valley companies to take risks and spend huge amounts of money on research. Maybe in the UK we need to accept that in order to play the game at the same level.

CASE STUDY

Applying cutting-edge techniques to impact on public good

Tom Smith is Director of the Data Science Campus at the Office for National Statistics.

CAREER HISTORY

From summer job to government data leader, via academia and spin out

I have been programming since getting a ZX81 in 1981, and quickly discovered that if you could program, you could get quite well-paid summer jobs from school. So in the 1980s, I was running statistical algorithms on mainframe computers in my summer holidays.

As an undergraduate at Cambridge I did physics, chemistry and various flavours of maths, and there was a big computational element to the work. My third-year thesis was developing computational models for identifying how spiral-shaped galaxies would form. There was understandably a big drive to make best use of available technology and processing power, and the physics group had an astonishing interest in how you could use technology – so I'm not surprised that Tim Berners-Lee was working at CERN when he developed the model for the web.

My first job was working at the Oxford University Social Disadvantage Research Centre group, where we were looking at using administrative data from housing benefit systems to measure or estimate levels of poverty and deprivation. The Centre really pioneered the reuse of government administrative data in this way, and I learnt that you needed to get very deep down into the data itself, how it was stored and managed, and how it was generated.

Alongside that work with government data, I also got interested in the potential for using data in different ways. In the mid-90s, artificial intelligence was coming out of a relative slump, and was being reinvigorated by work on neural networks and other evolutionary and adaptive systems led by groups such as Sussex University, and people like Geoffrey Hinton who are now very well known in the AI and machine learning communities. So my Master's and PhD at Sussex were spent evolving neural network structures that were used to control autonomous robots with vision systems.

Working on undergraduate, master's, PhD and then in research roles was a privilege and a fantastic opportunity to go where the research and my interests take me. That time to work on a problem and really get to understand it deeply is something that is unique to academia.

However, following my PhD, I felt I wanted to get my teeth into something more applied again. I launched a commercial venture 'spin-out' from the Oxford University research team that I had continued to work for, to make better use of government data for targeting and evaluating the impact of resources from government – essentially working on 'data for public good'. Over a decade or so, we built up the team to work on projects with local and national public sector agencies in UK and abroad, culminating in developing the Index of Multiple Deprivation on behalf of the government – using the methodology and theory developed by the Oxford team.

I finally found a way of merging my interests in data for public good, and AI and machine learning, when the ONS started setting up the Data Science Campus. It is part of a big drive by ONS to step-up on the way that we use different data sources, and data science techniques, to understand the economy and society. We've just celebrated our 2nd birthday. I am very motivated by the problem space that I am working on and it is one of the deep excitements about my current role that you can apply cutting-edge techniques and look for what is really coming out of the 'lab' in order to make use of it in the real world and have an impact on public good.

Teams, management and leadership

I now spend more of my time leading data science teams than being a direct data scientist. What I bring is the technical experience of having done this kind of work over 25+ years, linked to understanding what the end-point should be – ie, what is useful to deliver, not just what is technically or academically interesting.

In data science, development and training opportunities are critical. One big strength that the public sector has is the way that the organisation and leadership think about the development of their teams, which goes beyond everything I have come across. Career development and opportunities are something that the public sector scores very highly on.

Within industry and the public sector you are aimed very much at products, services and policies that are the result of quite a cross-functional, multi-skilled team. It is not only single individuals that are celebrated, so you have a lot more scope to celebrate the team itself. That is not always the case in academia where one of the drivers is the way that the various academic assessment exercises work. Working in teams is something that industry and public sector can provide to academics. To reuse a phrase from the Government Digital Service (GDS) when it launched: the unit of delivery is the team. This is not about individual heroes or unicorns.

One of the things that I have learned over the years is the importance of developing strong partnerships with people outside the organisation that you work with. In public sector this is critical to delivery, as there are often many groups involved in any piece of work. This comes down to understanding what the missions and values of other organisations that you might be collaborating, competing or partnering with are and how they align with yours.

Working across sectors and building trust

At senior, decision maker level, I would like to see better understanding of the potential and limitations of AI, data science and other technologies. I think we have made good progress, but it is still an uneven playing field. As William Gibson says, the future is already here, it's just not very evenly distributed. So I think we need to do much more around supporting and educating decision makers on the potential, and also the limitations of AI, data science and related technologies.

The public sector probably needs to shout louder about the important challenges we are working on, and to get better at asking for help from across different industries, sectors and organisations. For example there are skills and experience that we will sometimes need for short periods of time that we can use to strengthen our own delivery, and so could offer secondment opportunities to data scientists from academia and from the commercial sector.

Data science ethics is incredibly important. The models that data scientists develop, and the applications that we are working on, have such great potential that it has brought into sharp focus the importance of understanding the impact of our work. I think it is absolutely critical that the public sector and universities are transparent and open about what they are doing in this space while using data for public good, because there is an important point at stake around public trust in data use, and we have seen some high-profile examples from other sectors of how data can be misused.



CASE STUDY

Fostering entrepreneurship and approaching commercialisation

Ilya Zheludev is the Chief Data Officer for Jasmine22, a new technology venture being built by HSBC based in Hong Kong.

CAREER HISTORY

A focus on financial services, to change lives

I did a master's in engineering and business finance, a master's in transport and business management, a master's in financial computing, and a PhD in financial computing. During that time, I studied at UCL, LSE, LBS for a few courses, and Imperial College. My career has been predominantly in the financial services. I have worked in the front office of a bank and a macro investment fund. I have also worked in the IT side of an international multi service bank. I have worked in a fintech in the quant team. I have also worked in two investment banks in the front office.

I consider myself to be someone who is focused on technology and innovation in the data space, of which data science is a subset. I am motivated by the positive impact that use of data in a safe, confined, reusable, scalable way can have on day to day life. I think this will introduce the next shift of efficiency – similar to how the invention of computers transformed our access to knowledge. In my view, the societal efficiency shifts from data analytics are going to be fairly significant.

In academia, I had previously experienced a kind of wall in how to build products that touch customers directly. The work I am involved in now means that what I do can impact many people – although it is almost impossible to trace it back to me specifically. Being able to work on products that change people's lives is actually quite satisfying and I do not think I would have had that opportunity early on in academia. You can see the economy change and know that somewhere, at some point, you were a small part of that.

When I left academia, a data scientist was someone who was able to do the entire life cycle of a data challenge. That is: from collection of data, to storing it, then processing it, then making it available for analysis, then building models or inferences from it, and possibly even to deploying the models. I found the most valuable part of my academic life was working through this end-to-end cycle, rather than studying it, because by doing it you set yourselves challenges and you have to solve them. There is no option other than to solve them. No one is going to do it for you.

I made the transition to industry at a stage when I felt converting my experiences from academia could result in the greatest potential for customer impact. It was at a time when people with a data background at a post-doctorate level were starting to work in industries where commercial R&D in data science was emerging as a new priority.

Initiative and applied experience, and understanding the culture

This transition to industry brought me to projects where a combination of an academic and commercial focus was necessary to solve exciting, forward-thinking challenges. For me, achieving this blend of skills came from hands-on problem solving from an early stage. Being educated on principles is one thing, but I personally found that I learnt a lot more by doing. This learning can be from anything. It can be from work experience. It could be spending your free time trying to work on technical problems, or even just ‘playing’ with technical toolkits. It could be as part of a university degree, a school course, anything. The distinction I am trying to make is the difference between learning the theory and applying it to solve a problem. I would even go as far as saying that hands-on experience can be the most significant differentiator when it comes to a data science career. For example, if I am looking at junior data analyst CVs, the individual who has gone out of their way to work on problems in their spare time, or perhaps during some kind of placement or side project, really stands out. First, it demonstrates initiative and a direct interest; and, second, it demonstrates an aptitude to learn hands-on. I think that is key, because the environment you are thrown in to after university will require understanding a domain you have almost certainly not experienced to the same depth before, so the skill of learning while doing is extremely valuable.

I recall a course that I took on the side during one of my earlier degrees. It was ultimately on project management, but it was taught from the perspective of people from industry who had returned to academia. I found that highly relevant, and very important, because at the time I had almost no experience in the realities of implementing technology or data capabilities in a commercial environment. The best people to teach on topic like that are people who have done it themselves.

There are cultural nuances of different institutions in industry that are very difficult to understand before you join them. In my experience, the cultural element of an institution is not commonly translatable between organisations, but it can be enormously important on a day to day level. Whereas I found that there are quite a lot of cultural similarities between universities – academia is a kind of unified culture across academic institutions around the world. You can go from one university to another and feel relatively comfortable in a short period of time.

Working in industry, I can see there are technologies that are taking capabilities for which you would previously need a PhD and turning it into something that is a “point-and-click”. This is bringing data analytics to a broader audience, which is excellent to witness. I think that is an important step, because it means the accessibility to commercial data related roles is expanding beyond the need for higher tier qualifications. Technology is lowering the threshold or the barriers to entry.

Entrepreneurship, commercialisation, data sharing, and new training models

I don't think the British education system promotes entrepreneurship enough. Instead, it prioritises training for employment. But I think the economic landscape is changing due to the availability of data and tools to work with it. So today the barriers to entry for data-driven entrepreneurship are significantly lower than they were even a few years ago. It is a career route that I think we should be doing more to actively promote. It should be done carefully, with a clear statement of the risks – but something worth being aware of from a younger age.

If the UK government is interested in furthering the narrative of commercialisation of intellectual property (IP), it is a natural progression that individuals working in academia could be encouraged to participate actively in this. It is beneficial to the economy, and I think there should be more government support for organic transitions from academia to industry. This should be a strategic objective for a country, and so having relevant resources for it is important. For example, there could be more training-up to PhD-level of individuals who are open about intending to work in industry rather than academia. I think not being too prescriptive and rather building frameworks that allow people to explore paths that are most sensible to them is the right approach.

Data access is crucial and something I really struggled with in academia. It is not enough to just say, 'Here are dummy datasets', or synthesised datasets, or representative datasets. You want the real stuff. Mechanisms that enable that sharing across industry and academia, in a legal, ethical and compliant way are really important. Firstly, it is about collaboration. Secondly, it is about working on problem statements that are based on proprietary data. 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 and should have more involvement here.

On the flip side though, I think industry should play a bigger role in the training of people, regardless of what career stage they are at. I can see it being a lot more collaborative than the model of 'Here is a job. You can apply for it'. Instead, 'We are helping to upskill people, expose them to our company, give them training, give them some grounding and capabilities around the problems that the company faces', is arguably a more compelling and sustainable recruitment path. I also think it can be across the entire education spectrum, and I do not see why a company cannot offer training capabilities to anyone, regardless of age. Ultimately the pace of technology evolution is ever increasing. We should all be doing everything we can to enable people to be a part of that change.

The importance of continued learning

Dr Maria Wolters is an Alan Turing Fellow and Reader in Design Informatics at the University of Edinburgh.

CAREER HISTORY

Academia to industry and then back

My official title is Reader in Design Informatics and I am a programme director for two very closely related MSc programmes. I teach and do research on eHealth, health informatics, human computer interaction and accessibility. I work with data, but I am more of a communication and speech scientist than a data scientist.

I graduated with a master's in computer science from the University of Bonn in late 1997, and then continued to work there as lecturer while I did my PhD. Then I switched to industry for three years, but it turned out that it was not what I wanted to do. However, I stayed because there is a rule that in Germany that if you have at least three years' industry experience you have another academic route open to you, a professorship at a University for Applied Science.

I work on what makes information technology accessible to older people, what makes people use or abandon any sort of tracker that they use to log aspects of their health, and how we can meaningfully interpret the processes with which people generate health data.

For me, data is not the be-all-and-end-all. I use it to study human communication and human-computer interaction, so it has a distinctly utilitarian role.

Improve research culture in academia

I worked part-time as a Research Fellow for quite a while after re-joining academia because I then had two children. This meant that by the time I had settled into my current career, I was no longer eligible for a lot of early career or even mid-career scholarships.

We need different job descriptions in academia. Unfortunately, support roles are being cut or are no longer being made permanent. For example, there are very few permanent research programmers; they need to be funded project by project and that creates a lot of insecurity.

Upskilling and retraining throughout your career

Rather than a focus on developing skills, we need to work on a framework where people can regularly take time out of their work to learn, such as short courses. The field is moving so rapidly, especially now with deep learning. Suddenly, there is this whole explosion in new methods. So instead of saying, 'Look, these are skills that you have to have', it is more about learning to learn and setting aside time to teach yourself new skills.

This could be achieved through something like a continuous professional development model that the health profession has. I wonder whether this could be brought back into a business, to have a group of people who take a MOOC together or have a network or a database of courses. I know that lots of people already take Coursera courses outside of their normal working hours to upskill.

That would be useful because it is not just the theory that these courses teach you; but also the practical tools like Hadoop or TensorFlow. It is about building up expertise. Plus, you need to be continuously updating your skills as the next tool comes in.

For future upskilling, I think that Post-docs and PhD students should be targeted because they are hit by academic precariousness. There is a need to offer a wider range of training to them so that they have a wider range of options open to them.



CASE STUDY

Rethink UK university teaching to produce tech-ready graduates at scale

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

CAREER HISTORY

Academia to Industry and back again

I studied computer science in Cambridge. During my doctoral studies I had the opportunity to go and spend some time in the United States and I did a three-month internship at AT&T labs. I finished up my PhD and went into a postdoctoral position under the supervision of someone I worked with at AT&T. That led into permanent positions at Bell Labs and then at AT&T again for about six years before moving to academia, coming back to the UK. I am aware that I have been very privileged in my journey.

Working at AT&T was particularly critical for me being able to think of myself as working within data science. It was a private research organisation within a much larger company that was not strictly organised along traditional disciplinary boundaries. There were seminars, talks, conversations in the canteens and conversations in the corridor. Within that organisation there were computer scientists, mathematicians, statisticians and people who may have been more from engineering or physical sciences backgrounds. There were people working in data visualisation, and others in photonics thinking about signalling. It was an environment where it was so perfectly normal to be interacting with people whose education and background was quite different to yours that you did not get stuck into a silo.

KEY TO SUCCESS

Build a research portfolio

I can give a variety of reasons why I made the moves I did. My now wife was based in the UK; I was living in America at the time, so for family reasons it made sense to move. Another factor was that AT&T was a great work environment to be in, but it was a research organisation within a larger company and it was therefore at the mercy of the ups and downs of the parent company.

What made it feasible for me to move from industry to academia was 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 published rather than kept as a corporate secret.

In some sense, although at various points in my career I have been notionally in industry or academia, I have essentially been able to maintain an academic profile throughout. Were that not the case, I think it would have been much harder to make that transition because I would not have had the profile that academia expects. 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 they will say, 'Before we get started, can you do some programming exercises?' For someone who has been working on more theoretical topics, it can actually be a bit of a wrench to reorient to those more hands on expectations that organisations have.

Building foundational skills, rethinking university teaching and fostering a diverse community

There is definitely a need to develop foundational mathematical skills, including the ability to reason about linear algebra and mathematical representations of data, in particular.

What is more nebulous is the skill of having ‘common sense for data.’ This is the ability to look at quantitative information and draw conclusions from it. It may seem a soft or intuitive skill, but it could perhaps be formalised into something more precise. It is about being able to say, ‘something between the algorithm as I understand it conceptually and the behaviour of it, as measured empirically, is adrift, and I need to formulate and investigate hypotheses to explain this.’ Does the problem come from the data, the measurement or the experiment?

There are diversity issues across all stages of the pipeline. These go back all the way to reception year in school, I would say, and probably even earlier in our education system.

In the US, many students go into high tech or data science-relevant disciplines after graduating. However, at least locally within Warwick, it feels like the career paths have not moved so much in that direction. A lot of our students end up in the finance sector; a few of them go into computer games or education. I don’t see so many students leaving our computer science department and going into technology or data science.

The best conjecture I have is around some rigidity of thinking in the UK. We get our students to specialise at A-level at 16, then pick one of a limited number of degree options and have degrees that are actually relatively tightly defined. If someone comes to Warwick to do a computer science degree, they usually have to make a special request to be allowed to go outside the department to, say, take a module from the statistics department. It requires a fair amount of planning and commitment for a student to diversify their education at the degree level. It may not be reasonable to expect them to do so unaided.

To solve this, we can look to the US system where they have more fluidity to select options. The trade off is that students in the US system can come out of a four year degree having gone to a slightly lesser depth in any topic in order to meet their breadth requirements.

There are lots of structural issues about how you teach a topic that is so in demand at scale. Harvard runs an introduction to computer science and they have 1,000 students signed up for that one course every year. To service it, they have to do something radically different. They teach in the biggest lecture hall on campus with mic’ed up lecturers (and sometimes a DJ!). They have multiple levels of hierarchy of subgroup teaching and exercises. My view is that the traditional lecture teaching system scales up to 200 to 300 students with a single lecturer. Beyond that, you have to reconfigure: you have to do recording of lectures and maybe you do more MOOC style recorded lectures, but live exercises. Demand for data science may force universities to rethink teaching at scale.

Entry-level Analyst to Data Scientist at Government Communications Headquarters (GCHQ)

Aimee is Capability Researcher at GCHQ, a world-leading intelligence, cyber and security agency with a mission to keep the UK safe.

CAREER HISTORY

From academia to the security sector

I came to GCHQ right after university. I started off as an analyst but I quickly retrained, moving towards more of a data science role because there was such a demand for it. I would say I am a data scientist, although it is a very broad term and it seems to cover a whole lot of people. I focus on machine learning techniques and Big Data. There are other things which I do not do but I would consider a data scientist to do, like Hadoop or AI.

I came here because of the difficulty of finding a permanent job in the academic sector. I have friends in academia and they struggle to maintain funding. I have a guaranteed job after four years, whereas they all have very short-term contracts. Also, being here in GCHQ in particular, the opportunity to work on problems that help people was a big draw. I am answering interesting questions, learning new things, thinking carefully about problems and learning new maths.

I did not know what job I was going into before I started. The advantage of working here is that it is very flexible; I moved from being an analyst, which is less technical, to be a much more technical data scientist. But there are a lot of skills that I have not yet used, because I spend a lot of time getting the data in, cleaning it, going to meetings and talking to stakeholders.

As data science is ill-defined you have to draw on lots of different disciplines and work together. Nobody can know everything in the field as it is too big and moving too quickly. I work with software engineers, domain experts, mathematicians, statisticians and people who know about technologies that I do not. I spend a lot of time bringing together different teams and identifying ways that different partners can help each other. Being able to explain technical ideas in a way that non-technical people can understand is a key part. It is also important for our work to demonstrate real impact as we have to justify how we spend taxpayer money.

Open-minded approaches, upskilling on the go

I have learnt a lot about how to implement a result and interrogate it to work out where the flaws are. Not going through a data science degree might have been an advantage here. It is very easy in data science to say “this is how things are done”, and not necessarily question the assumptions behind it. I have benefited by being able to forge my own path.

It is much easier to upskill in data science than in other disciplines because there is so much available online, including Massive Online Open Courses (MOOCs). Thanks to these, people are catching up quickly. But there are some risks. Working in industry, every company wants a different set of skills from their data scientists, which makes it hard to define. People have heard of data science and know what it is about, but they do not necessarily know enough to ask good questions. To some extent, people think data is the answer to all their problems and so they throw a lot of problems at data scientists where it might be possible to solve it with a much simpler approach.

I think there is a shortage of in-depth knowledge of what the algorithms are doing and the ability to write new algorithms, which you might get on a computer science course. The software and database engineering skills to bring something from prototype to conclusion – so knowledge of Hadoop or other Big Data platforms – are also in short supply, where data science crosses over into software engineering.

Broader career advice, diversity in role models, and the importance of community

There should definitely be more diversity in role models. That would have been good for me, growing up, to see more diverse people going into STEM jobs. I think it is important at all stages of a career to have somebody like a role model. It shows you that “this is a career for people like me.”

It would be good to communicate at school that maths as an employment area is different from maths at school. I was not the ‘mathiest’ maths student ever. Knowing that it can be useful would have been good for me as there are so many careers involving maths. There is still the stereotype of maths being for ‘gangly boys’.

At work, we try to get people together to talk about the data science we are doing across different teams, to share techniques and read papers. We have a lot of people who are fairly new to data and so are trying to improve together and share thinking. We have had success with group Kaggle activities, which have always been quite fun. People generally like to work on problems together and learn from each other.

It is good to get an idea of what techniques are out there that you may not have heard of. Data science is changing so quickly that it is very easy to miss something great, like different Python libraries or workflows or techniques. It is easier to keep up with the field if you can hear how someone else has tried something, their method and how it worked for them.