Data Science Campus

Our first two years

2019

The power of data science for public good

Contents

Foreword

The Office for National Statistics's (ONS) Data Science Campus has a vital role in mobilising the power of data to help the UK make better decisions. At the launch of the new Campus in Newport in March 2017, I said that it would innovate with new methods and data sources providing opportunities to improve existing statistics and develop new outputs by working across government, industry, academia and charities in the UK and internationally. I also highlighted the wide range of training and learning programmes that the Campus would offer to build data science capability across the UK.
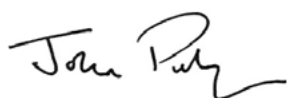
I have been greatly encouraged by what has been achieved so far in realising our vision for a new centre for the development and application of data science techniques. I am proud to introduce this first report on the progress made.

As well as recruiting experienced data scientists into the Campus, ONS has developed data science skills across the analytical professions throughout government. I was particularly proud to attend the graduation ceremony for the UK's first data analytics apprentices in Newport in November 2018. The eight graduates are joining a unique career pathway that will see them use cutting-edge tools and technologies to provide statistics and insights to help shape policy across the country.

Professor Sir Charles Bean's 2016 Review of Economic Statistics envisaged the Campus recruiting a cadre of data scientists along with active learning and experimentation facilitated through collaboration with relevant partners. The Campus has opened up many possibilities of working together with respected leaders in the field of data science, such as the Alan Turing Institute. These partnerships enable the Campus to develop research, innovate, and improve and exchange knowledge and skills.

Through its research programme, the Campus is actively helping to produce better statistics and is supporting better policy and operational decisions across a diverse span of the UK public sector and beyond, as the case studies in this report demonstrate.

I look forward to the Campus building on the tremendous achievements made so far and playing an increasingly pivotal role in driving forward the advancement of data science for public good.

John Pullinger
National Statistician

Introduction

It was a huge privilege for me to join ONS two years ago as Managing Director of the Data Science Campus. What a unique opportunity to put data science at the heart of decision making in the UK and help influence the most important policy issues facing the country.

We strive to achieve this in three ways:
- applying data science tools, methods and practices to strengthen statistics and evidence for policymaking
- innovating by assessing the value of new data sources and techniques
- improving data science skills across ONS and beyond

We have worked across government to deliver dozens of data science projects covering a broad span of topics including developing faster and more granular economic indicators, understanding trade, contributing to the public health debate and monitoring sustainable development. We have applied new data science tools, techniques and practices and investigated some of new types of data emerging from the data revolution.

Alongside data science delivery, we are growing capacity and supporting the data science community. We have a target to deliver 500 qualified data scientists for government by 2021 and, by working with ONS colleagues, the UK public sector and international statistics agencies we expect to exceed this. We have put in place a wide-ranging development programme from school to post-doctoral level which is attracting global interest.

Finally, how we work is important. By growing an experienced, diverse and creative data science group inside government we can demonstrate the value of government having direct access to world-class data science skills. By working in partnership with academia, industry and civil society organisations we can improve UK public sector access to data and data science skills. By developing a collaborative culture of working openly and supporting reuse of our work, we can maximise the impact of our programme.

I'm really proud of the great start we have made at the Campus. There is much more to be done. The next and critically important phase is to continue scaling the impact of data science across government. So as well as reviewing what we've achieved so far, the final section of this report looks forward to the Campus making a vital contribution to the challenges ahead. I can't wait!

Tom Smith
Managing Director
Data Science Campus

Who we are

The Data Science Campus (the Campus) is part of the Office for National Statistics (ONS), which is the government's National Statistical Institute and the UK's largest independent producer of official statistics. The Campus was created in response to the review of economic statistics published in 2016 by Professor Sir Charles Bean. The review recommended that ONS set up a national hub for data science to harness the power of big data to help Britain make better decisions and improve lives.

Across ONS, data is being mobilised to help Britain make better decisions and improve lives. Improvements in economic statistics, especially productivity, financial flows, prices and trade, mean that ONS is providing the statistics and data that decision-makers need for a modern economy. The upcoming 2021 census will be the first of its kind, "digital first", drawing on additional sources of information to create the most comprehensive picture of today's society. The creation of the Campus is part of this data-driven transformation of statistics in the UK.

The Campus launched in March 2017 with a core of well-qualified professionals, recruited mainly from industry and academia. Today, we have a team of nearly 70 experts actively delivering our ambitious research and academic programmes.

Our aim is to strengthen ONS and government expertise in data science across the UK, enabling quick, clear and relevant insight on the public issues at hand. Changes in society and technology have led to an explosion in data, making it more readily available and in richer and more complex forms. These developments mean that ONS has many opportunities to examine existing data and access new data sources and, through the work of the Campus, apply innovative statistical tools to these data sources to help with a better understanding of our society, our economy and our own lives.

What we do

As a response to these opportunities, ONS's Data Science Campus is delivering innovative research that will positively impact and enhance capability across the public sector using data science, machine learning and artificial intelligence. We are building new skills and applying new tools, methods and practices to support government decision-making and the UK Statistics Authority's Better Statistics, Better Decisions strategy.

Delivering data science projects

The Campus has delivered a series of high-profile data science projects that have provided valuable new insights for ONS and other stakeholders across government and beyond, making an impact on public policy over the last two years. Our programme spans the economy, trade, the environment and society. We publish technical reports on our projects and routinely make the code available for others to use via GitHub.

PARTNERSHIP IN ACTION

We worked with data scientists at the Department for International Trade (DIT) to analyse over half a million responses from their free trade agreement consultation data. We used a variety of advanced text mining and text summarisation techniques to provide insights that are being used by DIT policy teams to inform decision making, providing invaluable support for delivering free trade agreement outcomes in DIT and other government departments.

Luke Shaw
Data Scientist
Data Science
Campus

"Working at the Campus is a real privilege. The breadth of technical expertise and project variety, coupled with access to the right hardware and software, leads to genuinely exciting outputs for the public good."

The project process at the Campus allows teams to work flexibly while guarding against risks such as delays in accessing data and poor-quality data. A vital component of the project process is a review to ensure that all projects meet the stringent privacy and ethics standards set out by ONS. The agile nature of the project process also enables the Campus to mitigate the risks of projects becoming unfeasible or not progressing at pace to an impactful outcome for the stakeholder. We hold regular check-ups at Project Board meetings, and maintain regular contact with stakeholders.

Nia Jones
Statistician
Welsh Government

"The Data Science Campus seemed to relish the challenge of building a tool to help us analyse access to services, specifically public transport travel times to key services. Welsh Government developers are now well equipped to explore the application of this tool to our imminent update of the Welsh Index of Multiple Deprivation, and the Campus has been helpful and responsive throughout the process."

Building capability

With a diverse group of data scientists, lecturers and trainers, academic programme and partnership managers, the Knowledge Exchange team delivers a range of capacity-building activities that support the growth of data science across the public sector.

From data science and artificial intelligence training and mentoring programmes - developed and delivered by our in-house Data Science Faculty - to apprenticeships, graduate placements and a bespoke MSc programme developed with university partners, the team acts as a conduit for leading edge data science skills to enter the public sector.

Skevi Pericleous
Data Science Lecturer
Data Science Campus

"Data science can enhance decision-making in areas that will have an impact and improve people's lives. The Campus has a fantastic vision of establishing a data science community across government, and up-skilling analysts in the public sector. I feel extremely proud to be part of the team that is leading the way towards government transformation and the use of data science for public good!"

"The Campus is a state-of-the-art facility that will help us to attract and nurture the brightest talents to work on some of government's most complex issues."

John Manzoni
Chief Executive of the Civil Service and Cabinet Office Permanent Secretary

Alison Adams
Academic Manager
Data Science Campus

"I have worked at the Data Science Campus for a year now, and I can honestly say no two days are the same. In developing and running the data science and data analyst apprenticeships, I have learnt about data science and the fantastic career opportunities it can bring. I feel very lucky to be surrounded by people who are passionate about their profession and are always willing to share their expertise and knowledge."

Working together

We work with UK and international partners, drawing on their expertise and resources and sharing the benefits of our own education and research programmes. Our partnerships with industry, academia and international organisations were vital in jump-starting our operation, and we continue to draw on their valuable support as the Campus grows.

PARTNERSHIP IN ACTION

The Campus collaborated with HSBC and The Alan Turing Institute on The Turing-HSBC-ONS Economic Data Science Awards 2018. A programme of nine economic data science projects have been awarded a total of £750,000 in funding to combine world-leading science with the potential for high impact outside academia to improve our understanding of how the economy works.

Jane Crowe
Partnerships Manager
Data Science Campus

"It is inspiring to work with our many partners who are rising to a range of challenges and rewarding to know that, by working together, we can help them meet those challenges while learning a huge amount from them. I feel the Campus is in a unique position to share that learning widely"

We're making an impact

19 apprentices recruited. **30** data science research projects complete or in progress. **32** government employees on our bespoke MSc programme. Working with **United Nations** on a range of projects. On target to deliver **500 qualified data scientists** for government by **2021.** Projects support the government's **economic, trade, transport, health** and **social** policy making. Partnered with the private sector on a range of **non-commercial project**s focused on **public good** outcomes. **29** projects completed through mentoring programme. **224 civil servants** received work-related

data analytics training. **85** students put through intensive **CPD courses**. First cohort of **apprentices** have secured **data science jobs** in government. Set up partnerships with **Centres for Doctoral Training** in **AI, statistics** and **data science**. Campus-led public sector **data science skills audit** announced in the 2018 budget. Lead partner in a new research unit, **Economic Intelligence Wales. 222** young people engaged with **STEM outreach activities**. Collaboration with **national statistical institutes** across the world.
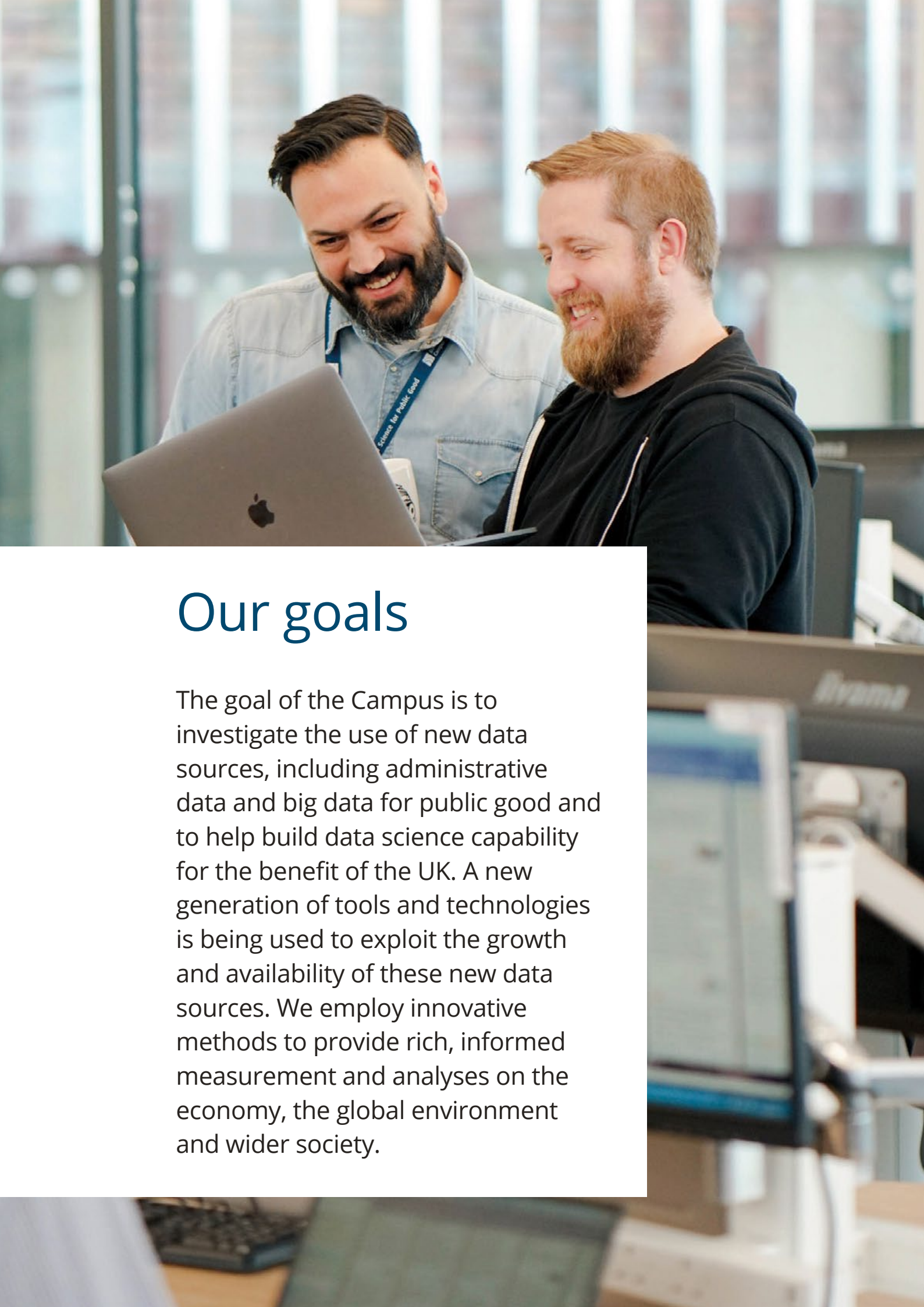
Our goals

The goal of the Campus is to investigate the use of new data sources, including administrative data and big data for public good and to help build data science capability for the benefit of the UK. A new generation of tools and technologies is being used to exploit the growth and availability of these new data sources. We employ innovative methods to provide rich, informed measurement and analyses on the economy, the global environment and wider society.

Our strategic objectives for 2019

Deliver research outputs

- Deliver 30 data science projects for UK public good.
- Publish case studies of new products and outputs used by ONS or elsewhere in government to provide greater insight to decision-makers.
- Set up new mechanisms to monitor the use and impact of Campus outputs in ONS and across government and record lessons learned across our projects.

Develop data science methods

- Publish new methods, code and analytical outputs for use by the wider community.
- Assess the value to ONS of non-traditional data sources and new technologies.

Grow data science skills in ONS and across government analytical professions

- Provide government with 150 qualified data scientists in 2019 (500 by 2021).
- Help departments strengthen their data science skills and understanding.
- Mentor projects on the Data Science Accelerator Programme or the ONS Data Science Academy.
- Support civil servants on the MSc in Data Analytics for Government and through our partners deliver continuous professional development (CPD) modules.

Form partnerships that provide access to new data and attract additional resources

- Increase the number of partnership agreements with universities, research institutes, international agencies and commercial businesses that lead to collaborative research or capability building activities.
- Harness commercial and other data sources and academic expertise to improve insight for public policy.
- Sponsor PhD and MSc students to carry out research in support of Campus objectives.
- Build data science leadership across statistics agencies to enhance knowledge sharing.

Our purpose and mission

We apply data science and build skills for public good across the UK and internationally.

We work at the frontier of data science and AI – building skills and applying tools, methods and practices – to create new understanding and improve decision-making for public good.

Our journey so far...

March 2016
Campus funding is approved

October 2016
First data scientist recruited, research commences

October 2016
Level 4 Data Analytics apprenticeship launches

December 2016
Sustainable Development Goals data visualisation platform launched

March 2017
Campus official launch, signing of first international Memorandum of Understanding with Statistics Netherlands

February 2018
ECLIPSE calorie counting project receives extensive front-page press coverage

July to September 2018
Art of the Possible training delivered to 560 attendees at Civil Service Live

September 2018
Campus project analysing unlabelled text from lorry manifests published

October 2018
Campus team reaches 50, with staff in Newport, London and Titchfield

October 2018
Project analysing public transport and access to public services for the Welsh Government published

October 2018
Campus outputs used to improve estimates of official statistics for trade

November 2018
Recruitment opens for Level 6 (Degree) apprenticeships

November 2018
"How green is your street" data visualisation platform published on ONS website

November 2018
Public consultation response analysis by the Campus used to inform international trade negotiations

March 2019
Degree level data science apprenticeship launched

March 2019
Rapid indicators for understanding the economy published

Source: Statistics Netherlands

Sofie de Broe
Head of Methodology, Scientific Director of the Centre for Big Data Statistics
Statistics Netherlands

"At Statistics Netherlands we were delighted to sign the first international partnership agreement in order to develop the field of data science in official statistics with the ONS Data Science Campus in March 2017. I have been really impressed with the pace and scale of progress made so far... ONS is now among the world's leading National Statistical Institutes in this field. We are looking forward to further intensive collaboration through staff exchanges, data science methodological research and capacity building."

Delivering data science

Data science is at the intersection of mathematics and statistics, computer science and domain-specific expertise. For the Campus, it's about improving our understanding of the UK's economy, communities and people, using novel data sources and techniques such as machine learning and natural language processing to better inform decision-making for the public good.

We use a range of innovative data sources, methods and approaches to deliver new outputs and products for ONS and elsewhere in government. Our projects span a range of functions of data science across diverse sectors and themes.

Main sources, approaches and themes of Data Science Campus projects

Better statistics and data

Support and enhance official statistical outputs at ONS by using new and existing data sources and finding better ways to exploit them.

Insight and analysis

Investigate data sources to help address policy challenges facing government. Sources could be new, existing or linked with other datasets to bring new insights to a government issue.

Operations and automation

Helping transform and improve the way government works in delivering techniques and capability to further the automation of systems and services.

Our approach to projects

Our ethos is simple: we want to work on projects that create the greatest public policy or delivery impact, or significantly improve our learning.
To prioritise project requests, we ask:

- does it add value?
- does it increase our understanding of the UK's economy and society?
- is it a stakeholder priority?
- what will we learn?
- can the learning be applied to other problems?

Deal breakers:

- is it ethical?
- is it possible?
- does it have an owner?

We follow a rigorous yet flexible process to deliver our projects, with the stakeholder and their desired outcomes always at the heart of that process. Wherever we can, we publish the findings and make the code available for others to use.
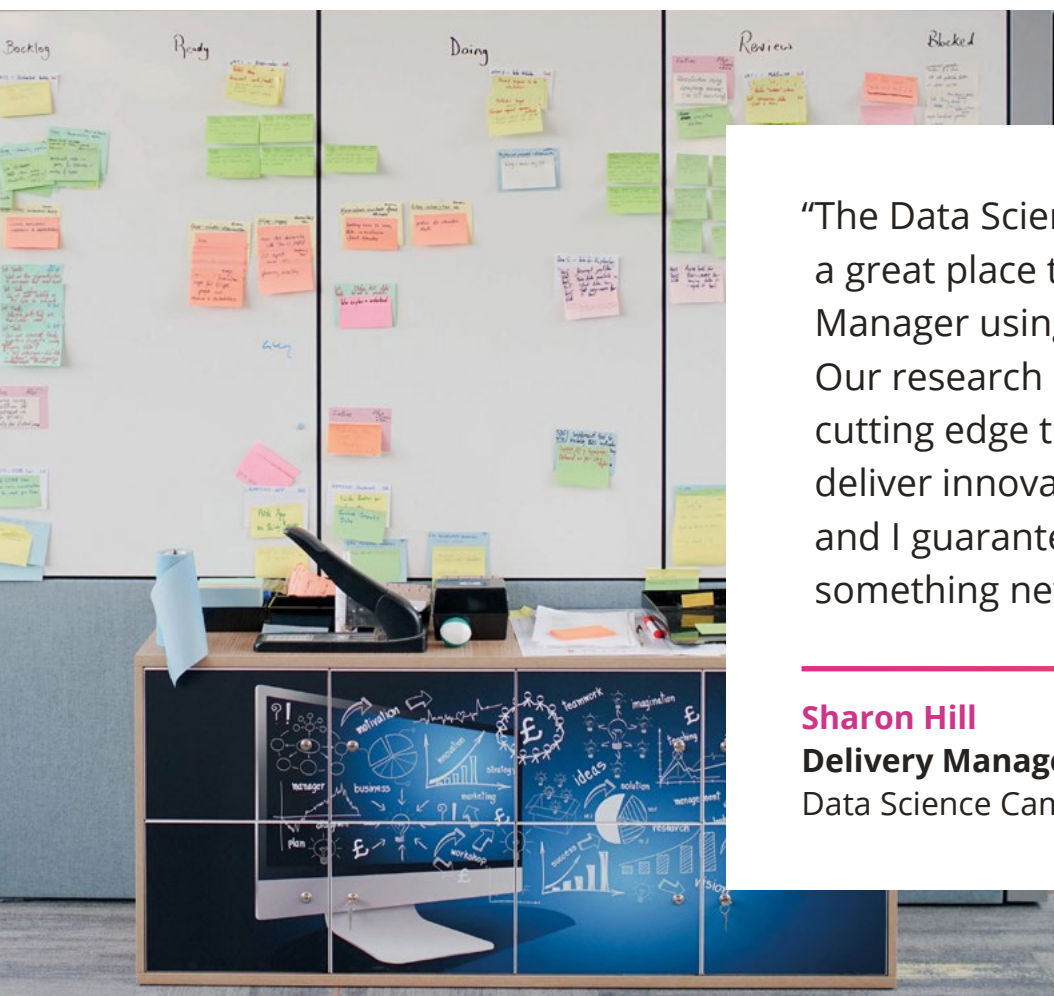
"The Data Science Campus is a great place to be a Delivery Manager using agile methods. Our research projects use cutting edge techniques to deliver innovative solutions and I guarantee, I learn something new every day!"

Sharon Hill
Delivery Manager
Data Science Campus

Figure 1: Project delivery process at the Data Science Campus

Ideas can come from within the Campus, within ONS or from external stakeholders.

Ideas that are successful in an initial pitch move to **backlog**. Projects are prioritised and approved by the Project Board.

Projects enter the **planning** phase when the data has been obtained and resources are allocated. The project plan will be defined and scoped.

Projects are **live** when they have full resource allocation. They are delivered using agile techniques. Stakeholder engagement is key.

Projects are in **roll-out** or **complete** when they are being handed over to stakeholders, and/or the code and report are shared through the appropriate channels.

Our projects

We are committed to developing world-class data science for public good. That means taking on challenging projects – finding and testing bold ideas, helping to shape government policy and having a real impact on those who need it most. And we're in demand. We're currently delivering on 26 research projects with four more in the planning phase, and as our team of data scientists continues to grow, we expect a continued acceleration in the rate of project completion and handover.

With a wide range of projects approaching completion, we are on course to meet our strategic goal of delivering on 30 data science projects by 2019. Tables 1 to 4 provide a list of our current and recent projects.

Table 1: Current and recent projects – Better statistics and data

Project	Goal	Stakeholders
Novel approaches to the Living Costs and Food Survey	Explore how we can apply computer vision and natural language processing techniques to the ONS Living Costs and Food Survey.	ONS Social Surveys
Sustainable Development Goals	Automatically identify bodies of water and analyse change over time, as well as explore access to all weather roads.	ONS Sustainable Development Goals
Classification of financial services	Explore classification of financial corporations to their detailed Standard Industry Classification 2007 using firm-level data.	ONS Economic Statistics Group

Project	Goal	Stakeholders
Mapping the urban forest	Assess the contribution of greenery in towns and cities to the UK's natural capital by creating a local-level dataset from classifying local street images and using image analysis and deep learning.	ONS Natural Capital
UK garden green space	Generate a better estimate of the green space within UK gardens, to improve the accuracy of ONS estimates of natural capital.	ONS Natural Capital
UN Global Platform – mapping the urban forest	Deploy our image processing pipeline used in the Mapping the urban forest project on to Algorithmia – a distributed computing environment used by the UN Global Platform project.	UN Global Platform
How green is your street?	Use vegetation index data produced by the Mapping the urban forest project to produce a data journalism and visualisation output, with the ONS Digital Publishing team.	ONS Digital Publishing
Payments data for public good	Work with Barclays to analyse payments data for public good.	Barclays, ONS Economic Statistics Group
Approaches for producing granular trade statistics	Develop a tool to support the production of more granular international trade in services (ITIS) output tables while meeting standards in disclosure control and accuracy.	ONS Economic Statistics Group

Payments data for public good

Use payments data to improve economic modelling

The Campus and Barclays are collaborating to investigate new ways of using payments data for public good, including analysis of the night-time economy and developing faster economic indicators to inform economic and monetary policy.

Project overview

Barclays and the Campus are collaborating to explore the rich potential of payments data for public good. We want to develop new and enhanced economic statistics taking advantage of the rich detail and timeliness of payments data. We are ensuring that we take into account privacy and ethical issues, by using only anonymised, aggregated statistics in which individuals' bank details cannot be identified.

Approach

The Campus hosted a knowledge sharing day with apprentices from Barclays and ONS. Following that, Barclays hosted a joint hackathon at their RISE London venue. The hackathon brought together 50 economists, developers, data scientists and analysts. The hackathon teams investigated ways to enhance or supplement economic statistics and the ideas of the winning teams have been taken forward in an ongoing collaboration.

This collaboration has led to two important pilot projects.

The first project looks at the night-time economy, and what payments data categorised by region and time of day can tell us about the scale and diversity of economic activity taking place at night.

The second project focuses on developing faster economic indicators. Payments data generated and collected very quickly could feed into the creation of a leading indicator for economic health. Using techniques such as anomaly detection and predictive models, the joint ONS and Barclays team – including economists, statisticians and data scientists – hope to produce new and important indexes and indicators to feed into national and regional economic policy.

Impact

The project will enhance local authority and government decision-making in a number of ways. Bank and card transactions and financial data offer a rich source of information about the economy and the project will help timely regional economic indicators to be constructed, which are important for informing effective economic and monetary policy. In addition, understanding the night-time economy enhances local authority decision-making about the local economy.

Mapping the urban forest

Develop an experimental computer vision method for estimating the density of trees and vegetation present along road networks

The urban forest project developed a tool using artificial intelligence to detect trees and vegetation in Google StreetView images. Our work could be used to improve estimates of natural capital by ONS.

Project overview

We used artificial intelligence to create a tool to detect urban trees and vegetation on the streets of 112 major towns and cities in the UK. Urban trees provide a wide range of environmental, social and economic benefits, such as improving air quality, and are known to be associated with lower crime levels and greater community cohesion. Within ONS, the Natural Capital Accounts team wanted to create an inventory of natural capital across the UK with a focus on detecting urban street vegetation.

Approach

We used computer vision technology, typically deployed in the emerging field of self-driving cars, to create a tool to detect trees and vegetation along roads. We made use of the Google StreetView platform as a data source to acquire street-level imagery for all major towns and cities in Great Britain at 10-metre intervals. The tree detection algorithm gives a score to each image of the density of vegetation present, accurate to over 90%. The data source and cutting-edge algorithms used in this work showcased the Campus's ability to operate at the very forefront of data science and AI within ONS.

Impact

We demonstrated our ability to describe, in detail, the visual components of a city in high resolution, including building density, number of cars, bicycles, people, signage, street furniture and other objects that describe an urban scene. This is a highly interesting geospatial dataset that could be used for a range of public policy applications.

This work could be used in the urban analytics domain, to feed into natural capital estimates. There are numerous studies linking green space to various social, environmental and economic indicators. For example, exploring the relationship between green space and other factors, such as indicators of well-being, offers an exciting direction for future research.

In collaboration with ONS Digital Publishing, an online "How green is your street?" interactive tool has been published, allowing visitors to the ONS website to check postcodes in Cardiff and Newport and receive a percentage greenery on their street.

Table 2: Current and recent projects – Insight and analysis

Project	Goal	Stakeholders
Categorising contents of lorries in cross-border goods	Improve understanding of freight at UK ports by processing unlabelled list data that are collected manually in lorry manifests with no supplementary information to allow aggregation of data.	Department for Environment, Food and Rural Affairs
Public transport access to services phase 1	Enhance understanding of public access to important services by creating a tool using multimodal (private and public) transport networks.	Welsh Government
Public transport access to services phase 2	Assess service accessibility by deploying our initial access to services project for generic use across the UK.	Welsh Government
Risk factors for loneliness	Determine the risk factors for loneliness across the UK with good geography, using health data as an outcome measure of loneliness and treating loneliness as a hidden variable.	ONS Public Policy Analysis
Indicator for housing quality	Explore new sources of evidence for indicators of housing quality, for example, energy efficiency.	Welsh Government, ONS Sustainable Development Goals
National Materials Datahub	Investigate what data science methods could be used to create the "go-to" source of materials information in the UK, open for public good.	Department for Business, Energy and Industrial Strategy, Department for Environment, Food and Rural Affairs
Flows of tenants within social residential accommodation	Explore mathematical models to simulate tenant flows, and clustering techniques to represent the different patterns of support and care provided.	Hafod housing association, NHS Scotland, ONS Public Policy Analysis

Project	Goal	Stakeholders
Identifying emergent trends from patent data	Identify ground-breaking products and technologies by applying machine learning to patents and data on emerging technologies.	Department for Business, Energy and Industrial Strategy, Cabinet Office, Intellectual Property Office, ONS
Understanding characteristics of high-growth firms	Explore how non-traditional data sources such as geographical features and web-scraped data can be combined with more conventional business data to help understand the characteristics and behaviours of high-growth companies.	Department for Business, Energy and Industrial Strategy
Extracting economic signals from internet bandwidth consumption data	Explore if it is possible to extract economic signals and insights from publicly available internet bandwidth consumption data.	ONS Economic Statistics Group
Economic impact of the UK fishing industry on local areas	Publish an online, interactive tool for policymakers, using ONS data to produce local-level economic indicators and data visualisations for the fishing industry.	Department for Environment, Food and Rural Affairs, Scottish Government, Welsh Government
Analysis of Automatic Identification System (AIS) data to understand shipping and ports	Explore the operation, use and relationships between ports in the UK at a macro level and the behaviour and operational characteristics of ships at a micro level.	Maritime and Coastguard Agency, Department for International Trade
Evaluating calorie intake	Support public health policies by improving our understanding of how much the UK is eating.	ONS Health Analysis and Life Events

Department for Business, Energy and Industrial Strategy

Identifying emerging trends from patent data

Develop information retrieval software that can retrieve popular technical terminology from documents.

This project, which identified emerging trends in patent data, enabled the Cabinet Office and the Department for Business, Energy and Industrial Strategy (BEIS) to analyse the innovation landscape by analysing data from patent applications. It also used artificial intelligence to enable BEIS to better understand the progress of the Clean Growth Grand Challenge policy.

Project overview

Our data scientists developed an information retrieval tool to retrieve popular technical terminology from patent abstracts and applied a method to quantify this. Thousands of new patents are granted each year, with thousands more applications filed unsuccessfully. The Cabinet Office and the Department for Business, Energy and Industrial Strategy (BEIS) want to use the data entries from patent applications to analyse the innovation landscape in the UK and globally.

BEIS also asked the Campus to look at grant documents to help them understand what proportion of government grants given to businesses support the Clean Growth Grand Challenge – one of the four Grand Challenges from the Industrial Strategy.

Approach

We developed an open source tool using a popular information retrieval method in natural language processing to extract popular terminology and other useful information from large document collections, such as patent applications.

We were also able to apply artificial intelligence, information retrieval and word embedding analysis to classify a data set of 70,000 business grants, based on whether they did or did not support the Clean Growth Grand Challenge. Our classification tool resulted in an 87% accuracy rate, based on a manually classified sample of 250.

Impact

BEIS data scientists have installed the patent technical term extraction tool on their servers and we have delivered training to their analysts on how to use these solutions to help shape policy decisions. The same text extraction software was also used on the ONS results for the Civil Service People Survey to identify key trends and terms being used in free-text sections of the questionnaire.

The project has also enabled BEIS to better understand the progress of the Clean Growth Grand Challenge policy, and will enable the identification and classification of clean growth-related applications in line with the Industrial Strategy without the need for manual processing.

Department for Environment, Food and Rural Affairs

Categorising contents of lorries in cross-border goods flows

Explore how to process unlabelled list data that are collected manually in an uncontrolled manner.

This project established a method to process messy, unstructured data on the contents of lorries passing through ports, using natural language processing. By enabling products to be grouped into categories and subgroups, the output has enabled the Department for Environment, Food and Rural Affairs to gain insights that were previously unavailable.

Project overview

Our data scientists used natural language processing (NLP) and a pre-trained word-embedding model (FastText) to group item descriptions based on meanings of free-text words rather than syntactic similarity. Lorries travelling through UK ports provide short descriptions of their contents but entries are often written in the driver's shorthand and include misspellings or typographical errors. The Department for Environment, Food and Rural Affairs (DEFRA) wants to better understand UK trade on sea routes and the type of goods travelling through UK ports but these insights were not possible from the pre-existing unstructured data.

Approach

Using NLP, we developed a tool that groups items into categories such as chemicals, vehicles, food, building materials and metals. The model can even identify subgroups, such as classifying cars by country of manufacture.

We were then able to create a pipeline, which used an optimised version of FastText to cluster items, automatically generate cluster labels and organise the data into a hierarchical dataset with named clusters and subclusters.

Impact

The project has provided DEFRA with previously inaccessible insights into the movement of goods through ports, with analysis carried out on 12 major shipping routes around the UK to date. The next stage of the project is to work with DEFRA to evaluate the performance of the pipeline and apply it to other datasets, which will allow the Campus team to refine the methodology.

The Campus is making the tool open source and a number of other organisations have expressed an interest in using the method to apply to projects which involve similar free-text variables.

Analysis of Automatic Identification System data to understand shipping and ports

Explore the operation, use and relationships between ships and ports in the UK.

We analysed Automatic Identification System (AIS) data to understand and predict the movement of ships in and around UK ports. These insights could be used by stakeholders to analyse port statistics and model port relationships to predict delays and maximise efficiency.

Project overview

This [project](#) explores the operation, use and relationships between ports in the UK at a macro level and the behaviour and operational characteristics of ships at a micro level. A team at the Campus began this work in conjunction with Statistics Netherlands, while the Maritime and Coastguard Agency (MCA) provided the necessary Automatic Identification System (AIS) and Consolidated European Reporting System (CERS) data. Together, these two datasets provide frequent snapshots of the position, speed, heading, bearing and rate of turn for each ship, as well as details such as destination port and expected time of arrival for the voyage of each ship.

Approach

We developed functions to extract, decode, sort and filter AIS messages as well as use machine learning algorithms to classify the ships' moving behaviour and expected arrival times. These analyses allow the prediction of inbound delays and the modelling of the interactions between UK and international ports.

Impact

The maritime freight industry is of critical importance to the economic output of the UK, with almost half a billion tonnes of freight handled by UK ports in 2017, according to the Department for Transport. As the demands upon shipping freight are likely to increase in the future, a more in-depth understanding of the UK maritime shipping industry becomes increasingly important. Both the MCA and Department for International Trade will use the outputs of this project to:

- process big data containing location of ships and reports containing itinerary information
- analyse port statistics based on several criteria
- model port relationships between UK and international ports
- classify ship travelling behaviour and predict delayed arrivals of freight ships

The Department for Transport is reusing part of our code to decode AIS messages and generate port statistics that are currently produced by an external organisation.

In addition to these direct impacts, the Campus developed its ability to work with big data. We are applying the learning from this to work on projects developing faster economic indicators, including analysis of trade in goods and estimates of gross domestic product.

Evaluating Calorie Intake for Population Statistical Estimates (ECLIPSE)

Explore novel data sources to improve the accuracy of official statistics on calorie consumption from food.

The ECLIPSE project supports public health policy by improving our understanding of how much the UK is eating. We analysed the discrepancy between self-reported calorie intake data and actual calorie intake. Our findings were reported extensively on the front pages of several UK national newspapers and attracted mainstream broadcast media attention. The methods and code were replicated by Public Health England to analyse the latest available data.

Project overview

Project ECLIPSE (Evaluating Calorie Intake for Population Statistical Estimates) researched methods for improving estimates of the national population's energy consumption and explored data sources from within and outside of ONS.

An earlier report by the Behavioural Insights Team highlighted a disparity between self-reported calorie intake in official statistics and UK obesity levels. The report shows that calorie intake has decreased over time while obesity levels have risen.

Approach

Doubly labelled water (DLW) measures – a physical and far more accurate measure than reporting by individuals – were used to measure energy expenditure as an approximation of true calorie intake. The ECLIPSE project used this energy expenditure data from the National Diet and Nutrition Survey to understand the extent of the apparent reporting error by individuals and the factors that affect under-reporting.

Data scientists at the Campus found that the average under-reporting error for participants in the dataset was 32%. The DLW study also showed no statistical evidence of a decline over time in calorie consumption, supporting the conclusions of the Behavioural Insights Team.

Impact

This study raised interest around the issue of obesity and calorie intake, and received extensive and long-term press attention, featuring on the front pages of The Times and The Daily Telegraph newspapers, as well as receiving coverage on the BBC News website, Sky and The Guardian.

Collaboration with topic experts across government increased the impact of this project – we provided our code to Public Health England who have reused it to validate the findings on later datasets, and they were featured on the Today Programme on BBC Radio 4.

Our approach provides a practical solution for improving the accuracy of the calorie intake estimates at a national level, making use of an existing government data source.

Public transport access to services

Provide the Welsh Government with a new accessibility and travel time tool.

This project delivered new insights for social policy in Wales by developing a tool to analyse access to services using public transport. The tool uses open source data to calculate transport times to and from public services. The Welsh Government is exploring the potential of our tool to improve estimations for the Welsh Index of Multiple Deprivation.

Project overview

The public transport access to services project developed a new tool to calculate travel times to and from public services including the nearest food shop, pharmacy, post office and public library. The Welsh Government asked the Campus to create an application that would enable it to improve its estimations for the access to services component of the Welsh Index of Multiple Deprivation (WIMD), the official measurement of relative deprivation for small areas in Wales.

Approach

We used the open source route planner, OpenTripPlanner (OTP) to host public transport timetable data, and designed a tool in R to extract information from OTP about the route. We used the information to create informative visualisations that show the area that is accessible from a specific location within set timescales.

Impact

The Welsh Government is currently scoping our tool's use for improving WIMD measurements by increasing the frequency of estimations and real-time monitoring. Other initiatives by the Welsh Government such as the Valleys Taskforce programme, the Cadw website and mobile app, and the South Wales Metro programme could also benefit from this programme in the future by providing insights into public transport provision.

The next stage for this project is to open up the tool for online access by multiple users, including businesses and the public, to analyse location data.

Table 3: Current and recent projects – Operations and automation

Project	Goal	Stakeholders
Creating a business prices processing prototype system in Python	Develop a prototype for processing business prices in Python, rather than in proprietary systems.	ONS Economic Statistics Group
Automated report generation	Create a pipeline for automated report generation with access to online application programming interfaces (APIs).	Department for Exiting the European Union
Synthetic data using generative models	Create synthetic data using neural networks to enable safer data sharing between organisations, and augment incomplete data.	Department for Business, Energy and Industrial Strategy, ONS Methodology
Improving the ONS search engine	Investigate the challenges of searching the ONS website and make recommendations for improvement.	ONS Digital Publishing

Table 4: Current and recent projects – Other government projects

Project	Goal	Stakeholders
Data science support across government	Four further projects have provided vital support to ONS and other parts of government in areas such as reporting platforms, corporate analytics and developing economic indicators.	ONS, other government departments and Devolved Administrations

Department for Business, Energy and Industrial Strategy, ONS Methodology

Creating synthetic data

Propose methods that generate synthetic data to enable safer, easier and faster data sharing between organisations.

This project assessed traditional and non-traditional methods for creating synthetic data. High-quality synthetic data can be used to improve the speed and security with which data are shared between organisations and to increase privacy.

Project overview

We are using a range of traditional and non-traditional methods to [create synthetic data](#) to make data sharing quicker and more secure. Government organisations, businesses, academia and other decision-making bodies would like to exploit big data, but the organisations that collect this information are often unable to share the granular data due to its sensitive nature. In this project, we proposed methods that generate synthetic data to replace the raw data for the purposes of processing and analysis.

Approach

There are twin aims to this project. The first is to create high-quality synthetic data that closely resemble the real data and are a suitable substitute for processing and analysis. The second is to ensure privacy – the synthetic dataset must not contain any identifiable data.

Our data scientists working on this are using a range of traditional and non-traditional methods to create synthetic data. Traditional methods for synthesising data include synthetic minority over-sampling. Non-traditional methods include state-of-the-art algorithms such as generative adversarial networks (GANs), variational autoencoders (VAE) and autoregressive models.

Impact

The Campus is in discussions with other government departments, including the Department for Business, Energy and Industrial Strategy, on how our work on generating synthetic data could improve their processes. Overall, the project is contributing towards a safer, easier and faster way to share data between ONS and the research community in cases where the real data are sensitive.

Building a world-class knowledge centre

The ambition of the UK government is to have one of the most digitally skilled populations of civil servants in the world. The Campus plays an important role in this by building public sector data science and AI capability through a range of learning and development programmes. These are delivered both directly, and in collaboration with important partners in ONS, the Government Digital Service, industry and academia.

Our ambition is to tackle some of society's big issues and use data to deliver public good. Together with our partners at ONS and across government, we want to embed new skills and embrace new ways of thinking to create a skilled and interconnected data science community. We aim to recruit new generations of talent from all walks of life; involve partners, stakeholders and experts from across government, academia and industry. We want to think big, exchange knowledge, and encourage participation so we can bring our work to life for the whole data science community to share.

We are aiming high. Under the UK Statistics Authority Business Plan, the Campus is tasked with training 150 qualified data scientists for government by the end of 2019, and 500 by 2021. We are already well ahead of schedule, and with the recent recruitment of new data science trainers and lecturers, we expect this number to increase significantly. We remain committed to strengthening our multi-strand capability programme, and are ready to meet the needs of both analysts and dedicated data scientists within ONS and across government.

> "We want people with the curiosity to ask the big questions, and the ingenuity to help answer them."
>
> **Tom Smith**
> **Managing Director**
> Data Science Campus

We're on target to exceed our goal of **150** qualified analysts in 2019, with **32** students on the Masters in Data Analytics for Government, **44** students on intensive CPD courses, **29** mentoring projects completed, **13** apprentices completing or on the scheme and **224** civil servants receiving classroom-based training in 2018.

Since our inception, we have delivered **2,336** hours of learning and development and engaged a total of **1,127** people across our capability programme, received **1,897** seminar registrations, and partnered with **29** government departments and other public bodies. We are on track to deliver **500** data analysts for government by 2021.

Harmonised career pathway

The Campus is supporting the development of a harmonised career pathway for data scientists in government, in collaboration with the government Analysis Function and the Digital, Data and Technology Profession. The pathway is supported by a wide range of Campus-developed and delivered learning and development programmes that cater to four levels of practitioner skill, from Awareness through to Expert.

Available to staff across ONS, the wider civil service and the public sector, these programmes include:

- awareness training for senior civil servants
- apprenticeships
- classroom-based training
- mentoring opportunities
- a bespoke MSc programme
- a guest lecture series from national and international partners

Head of data science / lead data scientist

These roles provide leadership and direction across a programme of multidisciplinary data science projects, managing resources to ensure delivery. They are recognised as a strategic authority with technical expertise in cutting-edge techniques, defining the organisation's vision. They are a role model to other data scientists and champion adoption of best practice. They communicate with senior stakeholders and convince them of the strategic value of data science. They are champions for the use of data science across government

Senior data scientist

Senior data scientists are experienced data scientists who provide support and guidance to teams. They are recognised authorities on a number of data science specialisms within government, with some knowledge of cutting-edge techniques. They may work on projects of high political exposure, value or complexity. They engage with senior stakeholders and champion the value of data science. They line manage more junior colleagues. They communicate the value of data science to senior stakeholders.

Data scientist

Data scientists are proficient in data science. They have recognised technical ability in a number of data science specialisms and provide detailed technical advice on their area of expertise. They draw on other technical and analytical standards from across government and industry. They promote and present data science work both within and outside of the organisation. They engage with stakeholders to demonstrate the value of data science and propagate data science skills in other teams. They line manage and mentor junior data scientists and manage small project teams.

Junior / associate data scientist

Junior or Associate Data Scientists are responsible for aspects of existing data science projects, whilst gaining valuable hands-on experience. They are able to apply certain data science techniques and work to develop their technical ability. They adhere to the data science ethics framework. They work as part of a multidisciplinary team with data architects, data engineers, analysts and others and provide limited advice on data science projects within teams. They identify and communicate lessons learnt during projects and follow good practice. They clearly communicate the value of data science work to stakeholders.

Trainee data scientist / apprentice

Trainee Data Scientists and Apprentices are given experience of practical data science work under supervision from more senior colleagues. They move from a strong awareness of core data science skills of coding, machine learning and statistics to a more effective working knowledge and develop their understanding of how to apply data science to business problems

Data Science Faculty

Our Data Science Faculty manages a curriculum of classroom-based courses aimed at leaders, policy teams and frontline staff to demystify data science. We run courses at both the Awareness and Working levels of the harmonised career pathway.

We develop and deliver these courses in-house. Our future plan is to work with partners to build further courses aimed at the Expert level. The Campus Faculty also runs "train the trainer" programmes, one of which was successfully deployed at the Ministry of Justice to help them achieve a goal of 250 analysts trained in the commonly used statistical programming language R. To date, we have trained around 130 analysts at the Ministry of Justice.

Awareness level workshops

Art of the Possible — 1 to 2 hours
This short workshop gives an overview of the use of data science in government. It is designed to demonstrate the value of data science to non-technical staff through a series of examples from across government.

Total persons trained in Awareness-level workshops in 2018 — **686**

Working level workshops

Introduction to R — 12 hours
Introduction to Python — 12 hours
These courses are designed for people who are new to programming and the R or Python languages. It provides the basic skills to operate in R or Python. Following this course, students can perform basic data manipulation and visualisation.

Data science with R — 2 x 12 hours
Data science with Python — 2 x 12 hours
These modules are hands-on and focuses on the reflection, collection and preparation stages of the data science process. We teach how to import data from almost any format into R or Python, and how to transform messy datasets into tidy ones. Students explore techniques such as visualisation to prepare data for analysis.

Spark and distributed systems — 4 hours
ONS is currently migrating operations to a more advanced platform (Cloudera) that hosts a series of distributed technologies. This course gives a quick introduction to the basics of using the main technology for analysis (Spark) while touching on other important technologies such as Hadoop (HDFS).

Total persons trained in Working-level workshops in 2018 — **194**

> "This one-week training course has been very fruitful for me. At the end of the module, I understood more about big data, database management, and the role Python and R programming can play in analysing big datasets. Now I am feeling more confident while talking about any data science element. I also started learning more about Python programming. I encourage other staff to learn about data science. Thanks a lot to the ONS Data Science Campus and the International Development branch that helped me to do this module."

Jean-Claude Nyirimanzi
Data Science Foundations CPD attendee
National Institute of Statistics of Rwanda

Masters in Data Analytics for Government

The new Masters in Data Analytics for Government (MDataGov) is a collaborative project between the Data Science Campus, ONS Learning Academy and academic partners across the UK. Launched in October 2017, this flexible, part-time programme aims to build data science capability across government by equipping civil servants with an important set of skills required from a modern government data analyst.

Students can choose to complete the programme within two to five years depending on their personal circumstances. There are four compulsory modules (Data Science Foundations, Statistics in Government, Survey Fundamentals and Statistical Programming) and eight optional modules from a range of courses in statistics and data science.

Total persons registered on MDataGov in 2018 — **32**

Continuous professional development

Civil servants and the Campus's national and international partners can study modules as stand-alone (assessed or non-assessed) courses for continuing professional development (CPD). These modules form the basis of the Practitioner level training managed by the Campus.

Total persons completed a CPD course in 2018 — **85**

Soumaya Mauthoor
Data Scientist
Home Office

"I am part of the first Campus-funded cohort to undertake the MSc in Data Analytics for Government at University College London (UCL) and a data scientist at the Home Office... I try to take full advantage of the access the MSc grants me to experts, both at UCL and elsewhere... Furthermore, I hope to share and get feedback on my dissertation project with other members of the MSc cohort and the Data Science Campus. The Masters gives me more self-confidence to prioritise my training when discussing timelines with managers and to put myself forward for more senior positions when the opportunity arises."

Data Analytics and Data Science apprenticeships

With the ONS Learning Academy, the Campus has been instrumental in developing and delivering two levels of apprenticeships.

Level 4 Data Analytics apprenticeship

The Campus recruited its first intake of eight apprentices in October 2016, the first Level 4 Data Analytics apprentices in Wales. In collaboration with ALS Training, the apprenticeship is a two-year programme, equivalent to the first year of an undergraduate degree in data analytics. Two years later, all of our first cohort had graduated and secured roles in government as data scientists.

The Campus recruited a further five apprentices in September 2017. They have worked alongside our experienced data scientists and been involved with a variety of projects, and now all begun their work placements across ONS ahead of their graduation later this year.

The apprentices have been an integral part of the Campus. As well as making a valuable contribution to our projects, they have been active in our school outreach programme, working with our STEM initiative and promoting data science and apprenticeships across ONS and beyond.

"The highlight of the scheme for me has been my involvement in the Urban Forests project at many different levels... I felt as though I was a part of the project team and that my contributions made a real difference to the project... Being a core part of the team porting the project onto the UN Global Platform has been a real highlight, and hopefully we can see the pipeline reused in other countries."

Joe Peskett
Data Analytics apprentice
Office for National Statistics

Level 6 (Degree) Data Science apprenticeship

Trailblazers is an initiative run by the Institute for Apprenticeships and is made up of a group of employers that come together as the creators and early adopters of new apprenticeship standards. ONS and the Campus lead the Trailblazer Group in England with over 50 employers (public and private sector) and 20 universities.

As a result, we now offer a three-year Level 6 (Degree) Data Science apprenticeship. The Welsh Government agreed to incorporate this into their degree apprenticeship delivery pilot and the Campus, in partnership with Cardiff Metropolitan University, is the first adopter. Level 6 apprentices have the exciting opportunity to combine the study of data science theory at university with working at the Campus alongside our experienced data scientists.

We began recruiting for apprentices in October 2018 and expect new apprentices to start in March 2019. The Campus is recruiting four apprentices and the Welsh Government will recruit a further two. We hope that this close partnership will allow us to build a network of degree-level apprentices and guide them through their early career in data science.

Recognition for apprenticeships

The Campus and the ONS Learning Academy have worked hard over the last two years to raise the profile of apprenticeships both within and outside the organisation, with ONS now having over 100 apprentices in many different disciplines. Being nominated as a finalist in the Large Employer of the Year at the Wales Apprenticeship Awards was recognition of this.

Large Employer Finalists at the Wales Apprenticeship Awards 2018
ONS and the Data Science Campus

Project mentoring

We believe mentoring is one of the most valuable and effective development opportunities we offer. Analysts from across the public sector can sign up to a range of different mentoring options:

- Data Science Accelerator
- Data Science Academy
- External mentoring of other government departments

The Data Science Accelerator is a capability-building programme, which gives analysts from across the public sector the opportunity to develop their data science skills. It started in 2015, and is backed by the Government Digital Service (GDS), ONS and Government Office for Science and the Analysis Function. We have been the South West and Wales hub since 2016.

Participants work on a three-month data science project. Having this protected time is an important benefit of the programme. Participants commit to spending one day a week at the Campus working on their project. Each participant is assigned a dedicated mentor (an experienced data scientist) and also benefits from peer support from other participants in their cohort.

We also run a similar programme called the Data Science Academy exclusively for ONS staff. Finally, we provide mentoring to teams across the public sector. These tend to be more flexible, with timetables built around individual work commitments.

29 Total mentoring projects completed since 2016

Table 5 provides a list of the projects that have taken part in these mentoring schemes.

Table 5: Projects and departments taking part in the mentoring schemes

Data Science Accelerator	
Project title	**Department**
Automation of object detection from satellite imagery	UK Hydrographic Office
Forecasting the condition of the school estate	Department for Education
Reducing potential harm through improved risk profiling	NHS Wales
Use machine learning to match individually collected prices to web-scraped prices	ONS
Pathway mining and analysis for patient-level data	Public Health Wales
Developing a tool to maximise the use of Trafficmaster data in Welsh Government	Welsh Government
AIS-derived products for improved defence situational awareness	UK Hydrographic Office
Sounding selection tool	UK Hydrographic Office
Better analysis and dissemination of the annual June survey of agriculture in England	Department for Environment, Food and Rural Affairs
Using machine learning techniques in economic statistics to improve survey methodology results	ONS
The Welsh name strategy	Pembrokeshire Local Authority
Automated patent casework allocation	Intellectual Property Office
Organisations' engagement with Innovate UK	Innovate UK
Project MERTZ	Royal Air Force

Data Science Accelerator

Project title	Department
Understanding the relationship between social care data and Ofsted inspections	Ofsted
Understanding public perceptions of teaching: automated analysis of free text data from online	Department for Education
What just happened? Using natural language processing to summarise patient notes and save doctor time	NHS Wales Informatics Service
Beach composition classification	UK Hydrographic Office
Text mining for public research impact evidence	UK Research and Innovation

Data Science Academy

Project title	Department
Identifying holding companies of special purpose entities	ONS
Big data and visualisations for apportioned regional tax revenues	ONS
Propensity matching with clothing and formula effect	ONS
Improving the accessibility of statistics on specific crime types	ONS
Machine learning as an alternative estimation method for later period VAT returns	ONS
Investigating possible markers of wealth and income in postcode sectors	ONS
Automatic classification of individual consumption by purpose (COICOP)	ONS

External mentoring of other government departments	
Project title	**Department**
Probability of success for the Training for Success programme	Northern Ireland Statistics and Research Agency
Stroke patients and effective prescribing	Northern Ireland Statistics and Research Agency
Modelling student loan repayments	Government Actuary's Department

SPOTLIGHT

Katie Davidson enrolled on one of the earlier rounds of the Data Science Accelerator before ONS and the Campus became a regional hub. She subsequently received one of the first sponsorships by the Campus to complete an MSc in Data Science at Birkbeck University. During this time, she was promoted to Head of Data Science. We are delighted to say that we are now helping Katie up-skill her team and she now sits on the board of the cross-government Data Science Skills Working Group.

"The Government Data Science Partnership and ONS Data Science Campus made it possible for me to explore the world of data science and successfully develop my career in this area. I would not have been able to complete the MSc without their support. I have now gone full circle as a mentor on the Accelerator and have responsibility for developing data science in my department. The Campus is now helping me help others in my department pursue wider data science careers."

Katie Davidson
Head of Data Science
Department of Health and Social Care

DATA SCIENCE ACCELERATOR IN ACTION

Since 2016, the UK Hydrographic Office (UKHO) has sent four analysts on the ONS Data Science Accelerator programme, working on a range of projects. For example, one project developed a sounding selection tool designed to detect seabed changes. The work focused on creating a process to significantly reduce the manual element of selecting the correct soundings from a survey to chart. It included using different data sources to select the most relevant depths for mariners.

The first project mentored by the Data Science Accelerator programme at the Campus helped Catherine Seale from UKHO automate the identification of objects in the sea. This project has since been developed into a live system in use at UKHO. It detects objects visible in the ocean on satellite imagery, such as wind turbines or oil and gas platforms. The system was presented at the Government Digital Service Sprint 18 meeting in May 2018, and to date, has processed satellite imagery covering 881,280 square kilometres of ocean, uncovering 342 hazards that were unknown to UKHO.

UKHO also announced at Sprint 18 that it would become a new hub location for the Data Science Accelerator programme, with a focus on geospatial projects.

Leading the way in infrastructure

The Campus has recently created a dedicated network, isolated from the core ONS IT infrastructure. The Campus network spans two physical secure data centres, providing high-availability, resilience and security. Users can connect from both secure corporate laptops and off-network devices such as MacBooks and Microsoft Surface Pros.

The environment is suited to explorative and development work and data scientists benefit from less restricted internet access, local administrator rights and the ability to install software packages without restriction. Virtual machines can be rapidly deployed with the latest data science tools and software. Data scientists also have access to General-Purpose Graphics Processing Units (GPGPUs) for machine learning purposes allowing data to be processed far faster than using traditional computer methods.

The infrastructure provides users with ample compute resource (processing, memory and large-scale storage) enabling virtual machines to scale as projects demand. This removes the previous limitations that users experienced when working solely off their laptops. Adhering to ONS policies and standards, non-sensitive data are now stored centrally in highly available data stores providing a single point of truth, enabling data scientists to better collaborate on projects. Users also benefit from being able to connect to external cloud providers such as Microsoft Azure and Amazon Web Services while using the Campus network.

The Campus network also contains a training environment providing students with a mixture of Microsoft Windows and Linux virtual machines, which are used to deliver various courses such as Python, R, Git, Apache Spark and natural language processing.

Working with others

Our ambition from day one was to be at the forefront of public sector data science in the UK. This can only happen through our partnerships and the exchange of data science knowledge between government, industry and academic practitioners – both in the UK and abroad.

Knowledge exchanges increase the use and understanding of data science within ONS and wider government; they enable access to data, tools, approaches and techniques developed at the leading edge of UK and international research. They also allow insights and methodologies developed within the public sector to be shared for the betterment of data science and the UK public as a whole.

In December 2016, we announced our first UK partnership with the signing of a Memorandum of Understanding with the Alan Turing Institute. Since then, we have been working and signing agreements with a wide range of prospective university, international and commercial partners to explore opportunities for collaborative research and joint programmes that advance the state of data science within government and across the entire field.

We actively welcome partners from academia, government and industry who wish to help us meet the demands and challenges posed by the evolving economy, and work to push the boundaries of data science research within ONS and beyond.

Academic institutions

We have built an extensive network with academic partners throughout the UK and beyond, providing funding opportunities for MSc and PhD candidates and delivering joint research programmes with a wide range of national and international partners.

Already, our partnerships with universities have allowed us to research, innovate and inject additional capability into the field of data science. We have shared research and resources and collaborated on various continuous learning initiatives to expand and improve knowledge across the UK.

University partnership and collaboration

The Campus has undertaken collaborative research with universities from all corners of the UK. We sponsor PhD students and provide a range of challenging short-term projects for groups of PhD and MSc students. Examples include a project for MSc students at Manchester University on how to capture changes of opinion for different groups in Twitter feeds. We are particularly excited to support University of Warwick on the 2019 "Data Science for Social Good" Summer Fellowship through collaboration with the Alan Turing Institute.

SPOTLIGHT

Hankui Peng received her degree in statistics and MSc in statistical practice. The Campus is sponsoring her PhD that focuses on exploring space clustering with application to text data. This research has applications in diverse areas including product categorisation, fraud detection and sentiment analysis.

"My PhD project focuses on high-dimensional clustering problems. It has been a truly interesting and exciting experience to work at the Data Science Campus. I have gained a great deal of insight into how data science research can play such an important role in economic applications and on such a large scale."

Hankui Peng
PhD Student, Associate Lecturer
Lancaster University

Flexible learning

Our three-to six-month project-based paid internships have proved very popular. Students can choose projects that support their own field of research or they can work on existing projects undertaken by the Campus. Currently, six MSc and PhD students from different universities have worked on projects alongside our experienced data scientists. We also sponsor four PhD students through the Alan Turing Institute. Students have access to our data science resources and research projects and benefit from a six-month placement with us during their studies.

Knowledge exchange

Our knowledge sharing events and support of forums such as university "data dives" has gone from strength to strength in 2018. For example, we have:

- collaborated with university faculties such as SAMBa (Statistical Applied Mathematics, Bath) to give students the opportunity to work with our data scientists and develop their own research
- hosted a knowledge exchange event with Lancaster University to explore the latest research and ideas into time series
- delivered lectures to PhD students at the Alan Turing Institute to outline the work and projects we work on for the public good
- held data science showcases at University of Warwick and University College London for PhD students

SPOTLIGHT

Twice a year, SAMBa hold a week-long Integrative Think Tank workshop, where they invite non-mathematical partners to set high-level challenges that the students can formulate into mathematical problems and work together to identify routes to a solution. In 2018, the Campus provided two challenges, and a team of Campus data scientists who spent a week working with the SAMBa students, supporting and mentoring them.

Susie Douglas
SAMBa Centre Manager
University of Bath

"The ONS Data Science Campus participated in our eighth Integrative Think Tank... the strength of the partnership developed led to a further collaboration with the Ministry of Planning in Paraguay who are using data science techniques to determine effective distribution of social benefits. We are looking forward to continuing to work with the Campus through joint projects, student secondments, and future research scoping. Having the support of organisations such as the Campus is invaluable to the successful delivery of SAMBa."

The future looks bright

We are developing and strengthening our academic collaborations into 2019 and beyond through the support we have pledged to our existing and proposed ESPRC and UKRI Centres for Doctoral Training Centres in AI, Statistics and Data Science. We also value our membership of Doctoral Training Centres advisory boards, helping to steer the direction of the training, projects and opportunities that aid the development of the skills of PhD students.

Our partners

University College London • Oxford Brookes University • University of Southampton • University of Bath • Cardiff University • Royal Society • Institute of Coding • STEM Learning UK • STEM Cymru • Royal Statistical Society • Cardiff Metropolitan University • Lancaster University • University of Warwick • Urban Big Data Centre, Glasgow • Open Data Institute • Alan Turing Institute • Gower College Swansea • Nesta • NIESR • University of Exeter • University of South Wales • Manchester University • University of Bristol

University of Glasgow • University of Edinburgh • The Datalab • University of the West of England • Queen's University Belfast • Birkbeck, University of London • University of Portsmouth • University of Swansea • Consumer Data Research Centre • King's College London • Royal Holloway, University of London • Queen Mary, University of London • University of Sussex • University of Plymouth • University of Oxford • Imperial College London • University of Manchester • London School of Hygiene and Tropical Medicine

Government

The Campus is driving data science capability across government.
We have worked with the government Analysis Function and Digital, Data and Technology Profession to develop a harmonised career pathway for data scientists, working with important government departments to focus on the development of agreed standards in technical skills for different job grades, as well as a consistent approach to skills assessment in recruitment and progression.

Public sector data science audit

As part of the Government Data Science Partnership, we launched a government data science skills working group. We agreed to conduct a data science skills survey across central government, leading to a HM Treasury request to widen the scope to the wider public sector. This was included in the 2018 Budget Red Book. The first phase of this audit began in January 2019 and we expect to publish the final report at the 2019 Government Data Science Conference in the autumn. We hosted the 2018 Government Data Science Partnership Conference in February 2018 and the first Government Data Science Community meet-up in November 2018, with 80 attendees from across government and the public sector. Planning is now underway for the next meet-up and the 2019 Government Data Science Conference.

National Materials DataHub

We have carried out an initial scoping exercise for the potential for data science to inform a National Materials DataHub. A joint collaboration with the Department for Business, Energy and Industrial Strategy (BEIS), the external advisory board was chaired by Campus Managing Director Tom Smith and included attendees from industry partners. The meeting discussed options for the next phase of work by the cross-government virtual team led jointly by ONS, BEIS and the Department for Environment, Food and Rural Affairs.

Economic Intelligence Wales

The Campus has been a lead partner in the creation of a new research unit – Economic Intelligence Wales. This is a new collaboration between the Development Bank of Wales, Cardiff Business School and ONS and was formally launched by the Welsh Government's Cabinet Secretary for the Economy in June 2018. The new research unit has responsibility for collating and analysing data to create an independent, robust and reliable platform to inform timely policy and funding decisions. Data gaps will be identified and addressed as part of this process.

"I see Economic Intelligence Wales as a strategic national asset and a key mechanism that can help shape future policy development within the arena of access to finance."

Ken Skates
Welsh Cabinet Secretary for the Economy

Our partners

Welsh Government • Department for Business, Energy and Industrial Strategy • Department for Environment, Food and Rural Affairs • UK Hydrographic Office • Department for Education • NHS Wales • Public Health Wales • Pembrokeshire Local Authority • Innovate UK • Ofsted • NHS Scotland • NHS Digital • Department for Exiting the European Union • Cabinet Office • Intellectual Property Office • Scottish Government • Maritime and Coastguard Agency • Department for International Trade • Royal Air Force • NHS Wales Informatics Service • UK Research and Innovation • Northern Ireland Statistics and Research Agency • Government Actuary's Department • Department for Health and Social Care • Government Digital Service • Government Office for Science

HM Treasury • Development Bank of Wales • Northern Ireland Civil Service • Department for International Development • Department for Digital, Culture, Media and Sport • Home Office • Care Quality Commission • NHS • Ministry of Defence • Office of Rail and Road • Ministry of Housing, Communities and Local Government • Ministry of Justice • Foreign and Commonwealth Office • HM Revenue and Customs • Ofgem • Forestry Commission • Wales Audit Office • Department for Work and Pensions • Public Health England • National Institute for Health and Care Excellence • Ofcom • Defence Science and Technology Laboratory • Companies House • Cabinet Open Innovation Team • Ordnance Survey • Bank of England • National Audit Office • Office for Artificial Intelligence

Industry

We work with industry partners on a range of non-commercial activities focused on outcomes for public good. In 2018, we partnered with Barclays to explore the development of rapid regional economic indicators using payment data. This led to the secondment of ONS analysts into Barclays, where they were able to work with the rich source of payments data held by Barclays, and benefit from the expertise and specific knowledge of their staff.

"We're really excited to play a key role in helping to support a better understanding of UK economic trends and growth. The hackathon was a great event to harness the excitement and expertise created through our partnership with ONS, and the winning teams have shown tangible evidence that payments data can indeed be used for public good."

Jon Hussey
Chief Data Officer Cards & Payments
Barclays

Other recent industry partnerships include:
- collaborating with Deloitte to raise awareness of data science across the Northern Ireland Civil Service
- partnering with PwC on a range of activities including their #GreatWales campaign, which focused on the impact of digital on Wales's economy, public services and infrastructure

We also supported a range of cross-sector groups aiming to increase the use and application of data science skills across the UK. These groups include:
- Royal Society Dynamics of Data Science Group
- Alan Turing Institute's Data Skills Taskforce
- Institute of Coding

PARTNERSHIP IN ACTION

Glass AI is a large-scale artificial intelligence system that reads, interprets and monitors the open internet. The company is building a new research resource for social, economic and market analysis. So far, Glass AI has digitally mapped the UK economy, tracking any topic of interest across hundreds of millions of web pages, and over 1.5 million organisations. We partnered with Glass AI on a project aiming to understand the characteristics of high-growth companies using non-traditional data sources, to inform policy decisions on investment and employment. Glass AI supplied us with data on a random sample of 30,000 UK active companies, including descriptions, sector classifications, mentions, news articles, job adverts and biographies of staff published on the organisation's website. This partnership has allowed the Campus and ONS to understand more about the use of non-traditional data sources in modern statistics.
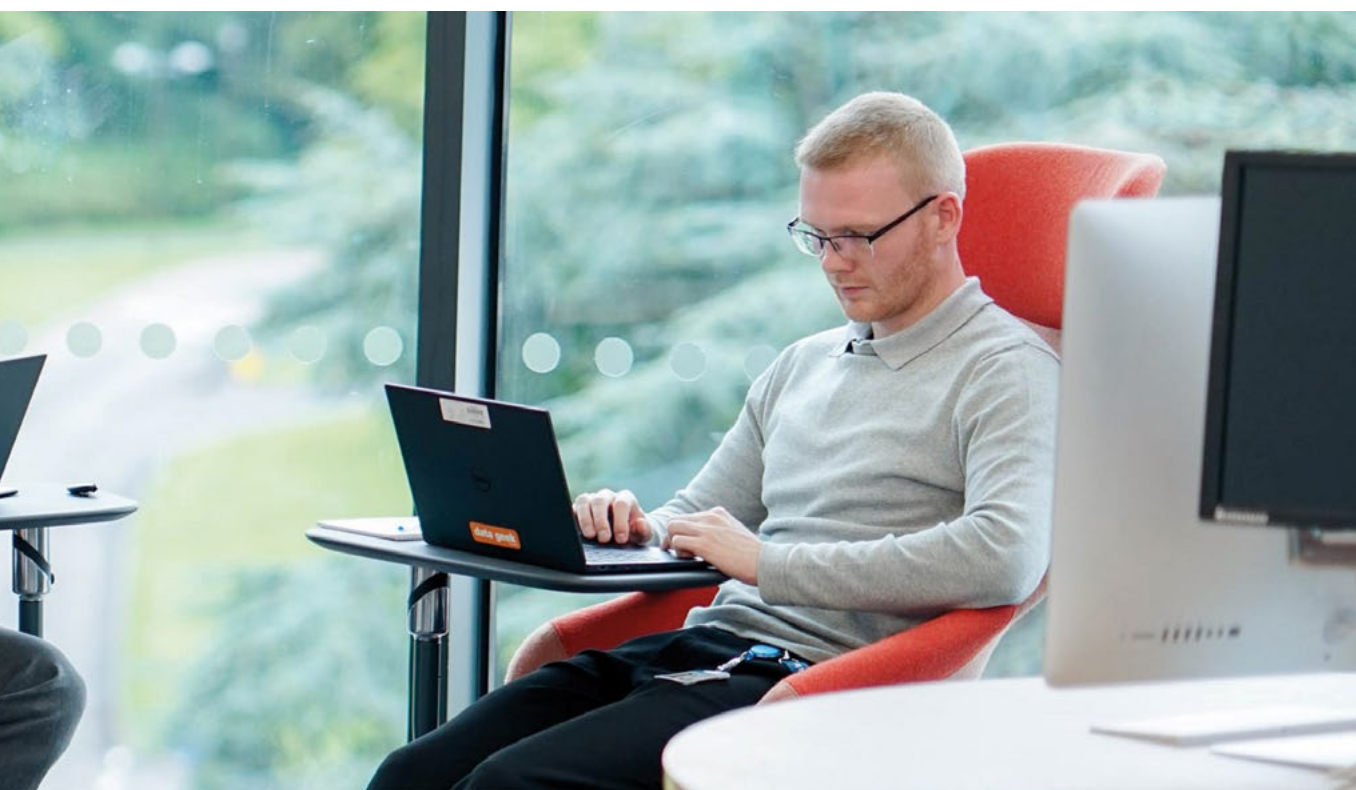
Our partners

Hafod Housing • Barclays • OpenStreetMap • GlassAI • Google • Deloitte • PwC • Cognizant • Valtech • Mango • Welsh Data Science Graduate Programme • EvolutionAI • NESTA • Wales Council for Voluntary Action • Hivemind • Data for Policy • The Behavioural Insights Team • Alacrity Foundation • Springboard

International

The world of statistics and data is constantly evolving and national statistical institutes (NSIs) from around the world are keen to hear more about the transformational journey of ONS and our own rapidly growing success story. We continue to be an active participant in the agreement between ONS and the Department for International Development (DFID) to support the modernisation of official statistics, initially in four African countries and with the UN Economic Commission for Africa (UNECA).

We regularly visit and receive visits from NSIs from across the world. In the last quarter of 2018 alone, the Campus welcomed delegates from China, the Republic of Korea, Singapore, New Zealand, Indonesia and Australia, as part of wider visits to ONS organised by ONS's International team. We have a formal memorandum of understanding with a number of these international institutions.

Rwanda's data revolution

ONS is supporting the Data Revolution for Rwanda initiative through the National Institute of Statistics of Rwanda (NISR). The Rwandan Government wants to build "an innovative data-enabled industry to harness rapid socio-economic development". Funded by DFID, the Campus has been providing strategic advice on the design and implementation of a sustainable and efficient data science capability plan.

Teams from the Campus have visited Rwanda on several occasions and collaborated with ONS colleagues to:

- provide advice on the legal, ethical and good practice aspects of data management and sharing
- advise on and support the creation of effective partnerships across the Rwandan public sector and academia
- provide advice on technical infrastructure
- assess skills requirements across the Rwanda Government, and provide training for 25 NISR staff
- the Campus is currently collaborating on two data science projects, one with NISR, and one jointly with NISR and the National Bank of Rwanda

Results are encouraging. During 2018, we saw Rwanda's Data Revolution policy taking shape and the building of the new Data Science and Training Centre.

United Nations

Data scientists at the Campus have been working with the UN Global Pulse Lab in Jakarta, a joint initiative of the United Nations and the Indonesian government. It is the first innovative lab of its kind in Asia. Pulse Lab Jakarta is working to close information gaps in the development and humanitarian sectors through the adoption of big data, real-time analytics and artificial intelligence.

The Campus has been helping the Lab to deploy several existing open source projects, including components of our Urban Forest image-processing pipeline, onto a platform in a format where they can be quickly adopted by other researchers in the international community.

The Campus is also contributing to the ONS work supporting the development of a regional data science campus for Africa at the UNECA headquarters in Addis Ababa, by facilitating a workshop on the use of data science within ONS.

International Monetary Fund

The Campus is leading on a joint International Monetary Fund and ONS project on mobile phone payments and remittances using commercial data. The project goal is to use anonymous peer-to-peer mobile phone money-transfers data to investigate the feasibility of producing economic and Sustainable Development Goal indicators. If successful we could develop a "tool-box" that potentially could be used by other countries with similar data infrastructure in the future.

"To take advantage of the opportunities from new sources of data, tools and processing power, NSIs are strengthening their use of data science; global groups such as the Data Science Campus and UN Global Platform are helping the international statistical community meet the challenge."

Ivo Havinga
Assistant Director of Economic Statistics,
Chief of Economic Statistics Branch
United Nations Statistics Division and Department of Economic and Social Affairs

Our partners

Statistics Canada • New Zealand Embassy • National Institute of Statistics and Geography (Mexico) • Brazilian Institute of Geography and Statistics • Statistics Norway • Eurostat • Health Quality & Safety Commission New Zealand • Statistics Korea • Statistics Indonesia • National Bureau of Statistics of China • Statistics Poland • United Nations Economic Commission for Africa

Statistics Netherlands • United Nations Global Platform • National Institute of Statistics of Rwanda • National Bank of Rwanda • National Institute of Statistics and Census of Argentina • United Nations Data Forum • World Bank • International Monetary Fund • Statistics Centre – Abu Dhabi • Australian Bureau of Statistics • Singapore Institute of Statistics

Governance

Harnessing the power of data science offers huge benefits for the UK government and the public at large. However, this is an emerging discipline and presents new challenges. Technology and statistical innovation are moving at pace and we need to constantly evolve our codes and ethics to match the highest standards demanded by the public sector.

We are in the business of harnessing the power of data to support the most important decisions facing the country, while ensuring that data are securely held and properly used.

Our governance supports these aims. The Campus is a Directorate within the UK's Office for National Statistics (ONS), which is itself the executive office of the UK Statistics Authority. The Authority is an independent body at arm's length from government, reporting directly to Parliament, with a statutory objective of promoting and safeguarding the production and publication of official statistics that serve the public good.

The Campus is led by Managing Director Tom Smith who reports to Heather Savory, Deputy National Statistician and Director General for Data Capability.

We have an Advisory Board that meets three times per year, and is chaired by the ONS Director General for Data Capability. The Advisory Board's main roles are to:

- provide advice on Data Science Campus activities and the delivery of its strategic objectives
- provide guidance on the development of the Campus and help the ONS executive give assurance to the Authority Board that the infrastructure is established and maintained in ways that serve the public good
- review how the Campus is working across ONS and government
- advise on the principles, policies and procedures of the Campus
- help resolve any high-level issues that inhibit the Campus achieving its goals
- help identify strategic risks to meeting Campus objectives and advise on their mitigation
- help oversee and guide public engagement and communications strategies
- advise on the opportunities for the development of the Campus.

Our Advisory Board members are:

Heather Savory
Deputy National Statistician and Director General for Data Capability at the Office for National Statistics

Dr Tom Smith
Managing Director at the Data Science Campus

Andy Shields
Director of Digital and Tech Policy at the Department of Digital, Culture, Media and Sport.

Glyn Jones
Chief Statistician for the Welsh Government.

Professor David Hand
Emeritus Professor of Mathematics and Senior Research Investigator at Imperial College, London.

Matthew Leach
Chief Executive Officer of Local Trust.

Professor Sofia Olhede
Professor of Statistics of Mathematics at University College London (UCL) and Director of the UCL Centre for Data Science.

Professor Piyushimita Thakuriah
Distinguished Professor and Dean of Edward J. Bloustein School of Planning and Public Policy at Rutgers University.

Caroline Bellamy
Chief Data Officer for the Ordnance Survey.

Claire Melamed
Executive Director of the Global Partnership for Sustainable Development Data.

John Taysom
Co-founder of Privitar, an enterprise software company.

Hetan Shah
Executive Director of the Royal Statistical Society.

Dr Sofie De Broe
Head of Methodology and Scientific Director of the Centre for Big Data statistics at Statistics Netherlands.

Professor Martin Weale
Professor of Economics at King's College London.

Geoff Little
Chief Executive of Bury Council

We follow policies and frameworks that operate across ONS to uphold ethical principles and safeguard data including new policies setting out how ONS looks after and uses data for public benefit, published in January 2019.

As with all ONS work, if we have a concern about the ethics of any of our research projects we consult the National Statistician's Data Ethics Advisory Committee (NSDEC). This committee was set up to provide independent and transparent ethical advice and to ensure that the use of data for research and statistical purposes is ethical and for the public good.

We helped the NSDEC develop a framework to help researchers, statisticians and data scientists assess the ethics of their research by scoring projects against six principles. This framework is now being used for the Campus's Research Programme to assess projects in their early stages. Any with a high-risk score are referred to NSDEC for a full ethical review.

Further information on the NSDEC data ethics self-assessment process is available via the UK Statistics Authority website.

"We are proud to have recruited a world-class team of data scientists from diverse backgrounds, including government, academia and industry, providing the Campus with a deep pool of skills and expertise."

Peter Fullerton
Deputy Director
Data Science Campus

Outreach and volunteering

We are involved with a range of outreach and charitable activities. These range from delivering data science awareness training to charitable organisations, to supporting outreach in schools and youth groups. Our staff also volunteer through Business In The Community and the Wales Council for Voluntary Action.

STEM Ambassadors

STEM Ambassadors are part of a national scheme led by STEM Learning, a partnership between government, charitable trusts and employers. The scheme enables volunteers to engage with young people and promote STEM (science, technology, engineering and mathematics) through career talks, mentoring, practical workshops and exhibitions. The Campus is proud to support several STEM Ambassadors.

ONS introduced a pilot programme in 2018 in partnership with STEM Cymru and the Engineering Education Scheme Wales. STEM Ambassadors from the Campus hosted groups of female school pupils for "Girls into STEM" workshops.

Our Ambassadors shared their energy and enthusiasm for STEM subjects with the visitors during a hands-on workshop, which included games to promote mathematics and data science as well as an introduction to coding session with our team of programmable Lego Boost robots.

We have 10 ambassadors made up of data scientists and data analytics apprentices. They have supported several engaging outreach activities in schools, clubs and within ONS. In all, our STEM outreach activity has engaged with a total of 222 young people and led to the award of the 2018 ONS Excellence Awards for Building Capability to the Campus.

While this pilot programme is now complete, in 2019 the Campus is looking to partner with bodies such as the Institute of Coding in Wales to identify ways to scale STEM engagement at a national level.
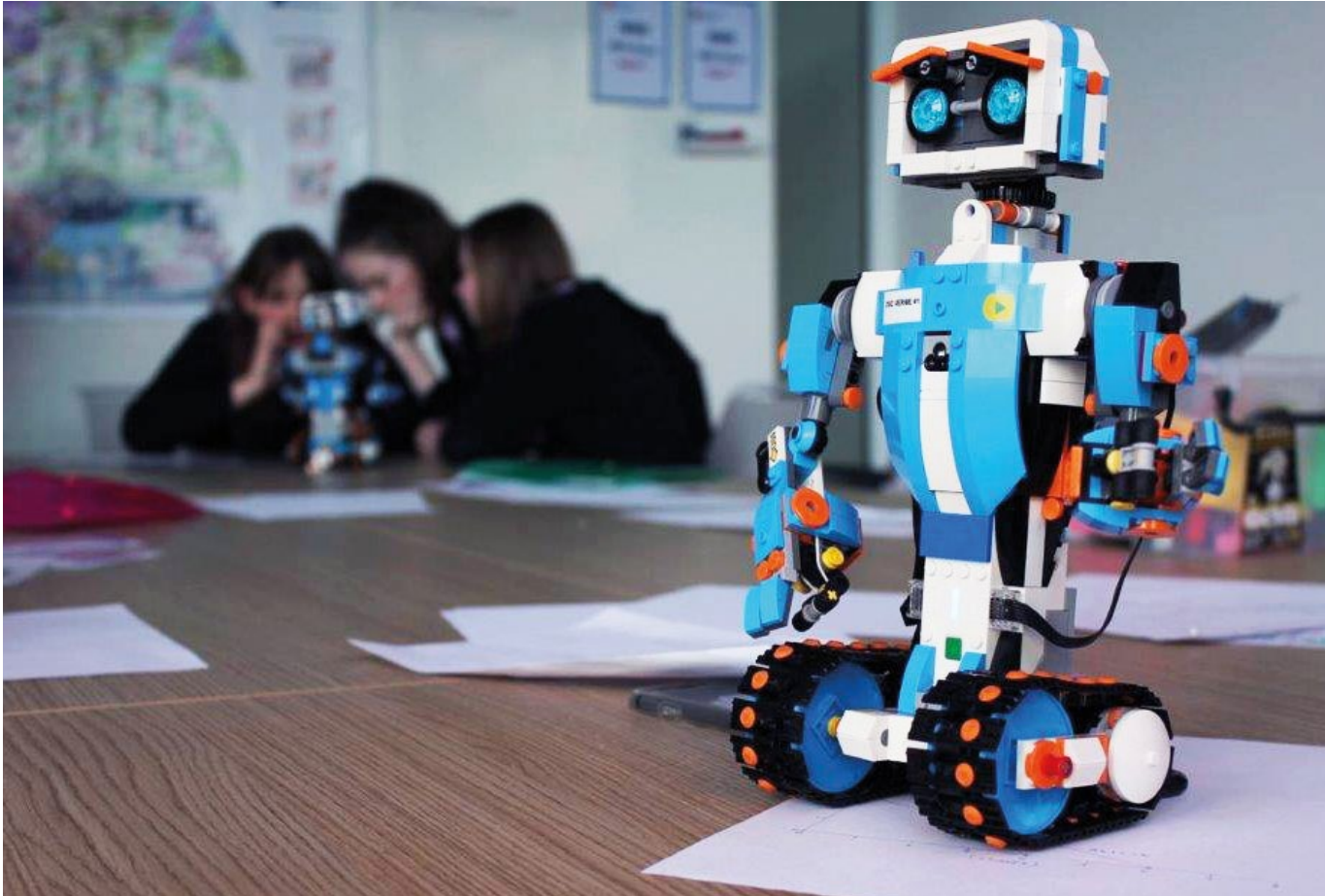
"Through our Girls into STEM workshops we wanted to encourage a passion for data science, leading to greater take-up of relevant subjects in school and at university... Demand for STEM skills is growing and it's important to address the low take-up of STEM subjects in girls beyond GCSE and to encourage girls to develop data science expertise."

Heather Savory
Deputy National Statistician and Director General
for Data Capability
Office for National Statistics

Volunteering

Through the outreach charity – Business In The Community – we have supported local organisations through our volunteering time.

Two Campus volunteering days were held in 2018:
- part of the Campus team spent a day at Global Gardens in Cardiff undertaking a range of gardening and manual activities – the Global Gardens Project is about bringing communities together with a vision to create a growing space that supports community-based sharing of food and cultures
- in November, Campus staff carried out internal and external building maintenance at Ystrad OAP Association, and socialised with the members over lunch – Ystrad OAP Association provides forums for those aged 60 years and over to meet up and plan activities

Data in the community

In early 2018, the Campus hosted a group of Welsh voluntary sector organisations. We showcased our range of projects and discussed opportunities to build data science skills across their community.

To encourage this sector to make better use of their data, we agreed to support the flagship conference Gofod 3. This is an event organised by Wales Council for Voluntary Action in collaboration with charity organisations across Wales. Over 400 delegates attended and our data scientists delivered a condensed version of "The Art of the Possible" to several conference delegates exploring how data science could relate to their charitable work.

Looking to the future

We've only just begun our journey but already we have achieved so much.

Two years on, what does the future hold for the ONS's Data Science Campus? The case studies in this report demonstrate how our early projects have delivered and we look forward to many more making a real impact for public good in the UK, providing new economic insights, using novel data sources to explore societal trends and assessing progress towards more sustainable development.

Over the past two years, we've built up a strong knowledge exchange team of academic managers, lecturers and trainers so that we can meet – and exceed – the target set by John Manzoni, Chief Executive of the Civil Service, to produce 500 data analysts across government trained in data science, by 2021. We've just recruited our first degree-level apprentices into the Campus and plan to extend ONS's apprenticeship programme to higher academic levels over the coming years.

We've seen our partnerships with academic institutions blossom and are looking forward to working with six of the recently announced Centres for Doctoral Training in mathematics, statistics and AI. Moving forward we will be exploring new collaborations, such as fellowships, to enhance the benefits to the country of our close links with academia.

We have barely scratched the surface of the potential to collaborate for public good with industry and we look forward to building on the success we have seen, for example, in our partnership with Barclaycard which is helping provide new perspectives on the UK economy.

Internally to ONS we will continue to work to embed data science skills across the organisation and to further the use of new types of data and analytical methods in the production of official statistics such as the Census and our economic outputs. Across government, we've already worked with most UK government departments and have created a significant hub of data science activity based in London. In March we reached agreement with the Department for International Development to establish a new Campus hub at their office in East Kilbride in Scotland, focused exclusively on international development. As we establish the Campus further, we expect to set up new Campus hubs in different parts of the country to ensure that the work we do is at the heart of UK public policy decisions – finding innovative ways to illuminate the unknown challenges ahead.

ONS is increasingly being recognised as among the front-runners of modernising statistical institutes worldwide, including in our ability to use big data and other novel data sources to improve national statistics. Overall we want to cement our place as a world leader in data science, working in collaboration with partners, including working through UN agencies and other bodies – developing partnership programmes that help countries across the world to build their data science capability.

With our talented and experienced Advisory Board now in place, we have a great launch pad so the Campus can help build on the achievements set out in this review of our first two years and help deliver a step change in the application of data science across the UK public sector. Above all, we want to help the UK public sector deliver the maximum benefit it can from the better use of data.

Heather Savory
Deputy National Statistician
and Director General for
Data Capability
Office for National Statistics

Tom Smith
Managing Director
Data Science Campus

How to work with us

For the last two years, we have been proud to work with, and support, our partners and colleagues across government, academia and industry – both in the UK and globally. We always welcome the opportunity to work with others to harness the power of data science to create new understanding and improve decision-making for public good.

We actively welcome partners from academia, government and industry who wish to join us as we seek to meet the demands and challenges posed by the evolving economy, and work to push the boundaries of data science research within ONS and beyond.

If you would like to collaborate with us or you are interested in joining our rapidly growing team, visit **www.datasciencecampus.ons.gov.uk** or email **datasciencecampus@ons.gov.uk**

Office for National Statistics

Contact us

Data Science Campus
Office for National Statistics
Government Buildings
Cardiff Road
Newport,
NP10 8XG

To find out more information about any of our projects or services
visit **www.datasciencecampus.ons.gov.uk**, email **datasciencecampus@ons.gov.uk**
or follow us on twitter **@DataSciCampus**