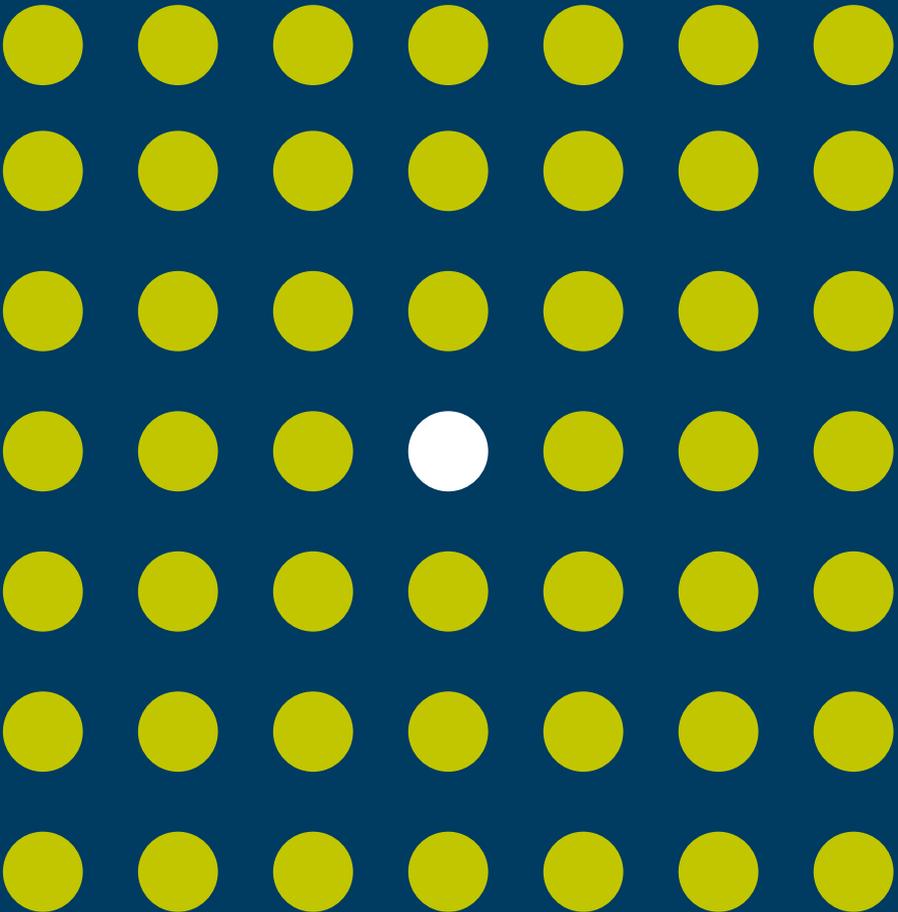


# Data Science Graduate Programme



**The data science  
graduate programme  
is an exciting opportunity  
for you to work at the  
heart of public sector  
data science.**

## **What will the Graduate Programme offer you?**

The digital revolution is changing everything, transforming how we work and play and how we view societies the world over. Management of data is becoming ever more important in a modern economy. Data Science is driving the development of better public services and improving the lives of citizens in the UK. In order to realise the potential that new data sources hold, we need to be alive to opportunities as they arise and ensure people have the right skills to unleash innovation through Data Science throughout the public sector.

## **Do you have the passion to work in this Data Science environment?**

If so, the graduate programme is a unique opportunity to work at the heart of Data Science in the public sector, to join experienced teams of Data Scientists and contribute towards the real world projects that they are working on. During the two year programme, you will be involved in collaboration and learn from colleagues across the public sector, and practices across government, academia and industry. You will be working on the biggest problems and challenges of our time, using your Data Science skills for the public good.



# The two-year Graduate Programme includes:

- Training in much valued data science skills and techniques
  - Project based learning with high profile public sector organisations including 10 Downing Street and HM Treasury
  - Collaboration and networking across the public sector
  - Professional development support including mentoring
  - Learning from the best in the public sector, industry and academia by attending data science conferences, seminars and events
  - A secure foundation to an exciting and rewarding career in public sector data science where you can use your skills for the public good
- 

# Graduate Programme structure

The first year of the programme consists of **10 training modules, spread over 12 months**. Average time commitment in a training month is 3 full working days which may include attending lectures and tutorials, working through independent learning or reading material, and/or assignments. Between these learning sessions **you will be working on projects in your home organisation** where you will apply your learning within a project context.

The first year of the programme allows you to spend structured time thinking about and practising data science. You will establish a **core set of data science skills** that complement the specific requirements of your UK government organisation. You will have the opportunity to **build collaborative relationships** with other graduates and leading UK government Data Scientists.

During the second year of the programme you will work on **specific projects that utilise your learning from the first year**. You will attend a series of events such as hackathons that encourage your Data Science skill development.

At the end of the programme, you will have **acquired new, highly sought-after analytical skills**, forming a secure foundation for a future career in public sector data science. You will have demonstrated an aptitude in the effective communication of valuable data insights, having engaged with a wide array of stakeholders. Crucially, **the Graduate Programme will have helped you to nurture your passion for cutting edge data science techniques**, equipping you for a job with purpose and a stimulating career.

# Curriculum Overview

The Graduate Programme curriculum provides a robust and exciting framework of learning materials designed to consolidate existing analytical skills. Our experienced lecturers work closely with graduate Data Scientists to continuously review and improve the content of our curriculum, ensuring our training materials are engaging, relevant and effective.

The aim is to provide you with a range of tools and experience in applying them, ensuring that you can make impartial, pragmatic analysis of the most appropriate software or technique in working towards given success criteria.

**Please note:** the current sequence of training is under review and is subject to change for 2021 to 2022.



# Year one

## **Module one: Data science foundations**

A welcome to the curriculum with an introduction to data ethics, data governance and data reproducibility. Introduction to programming is offered and you are expected to complement your programming skills by choosing the framework in which you have least experience at this stage, Python or R. Importing, basic processing and manipulation of data is covered, along with key syntax requirements, conditional flow and framework-specific data structures.

## **Module two: Designing effective workflows**

This module consolidates key programming skills for effective workflow management and improved reproducibility. Command line basics, introduction to Git and GitHub, and writing clean and efficient code introduce key programming skills required for effective development.

## **Module three: Statistics and visualisation**

Statistics in both R and Python are offered in this module, both courses providing the opportunity to implement core statistical techniques within a programming framework of their choice. You may choose to sit one or both courses. The data visualisation courses in R and Python prioritise best practice, effective visual communication and visualisation for exploratory data analysis.

### **Module four: Reproducible code and best practice**

Reproducible Open Science assists you in developing robust, well-integrated programs that achieve maximum benefit from programming software. Modular code helps you to consider the utility and structure of your code, ensuring programs are robust and easier to maintain. Unit testing introduces effective, isolated testing practices in the R and Python frameworks.

### **Module five: Machine learning**

Machine learning is introduced with an introductory course, taking a deep-dive into the theory and applications of widely applied machine learning techniques. Programmatic implementation of supervised and unsupervised learning techniques is then taught, using the Python programming framework.

### **Module six: Natural Language Processing (NLP)**

This module aims to build capacity in this much sought-after analytical suite of skills. Introduction to NLP will lay the foundation of text analytics within the Python programming framework. Intermediate NLP builds on this learning to develop resilience in basic/probabilistic language models and feature representation.



### **Module seven: Robust machine learning (ML) models**

This module explores evergreen and context-specific quality assurance measures to implement in machine learning models. Quality assurance and principles relating to predictive modelling / machine learning systems are introduced. The material will cover risks across the development cycle of ML models.

### **Module eight: Data science case studies**

Content dedicated to the practical application of data science techniques. Drawing on the expertise in the Data Science Campus projects and research teams, these courses will explore live and ongoing analyses that make use of cutting-edge techniques. You will explore project workflows, helping to embed theory in context and work on the application of skills to case studies.

### **Module nine: Databases & Big Data**

This module's aim is to equip you with the skills required to apply your prior learning to big data. This includes querying databases to return the required data subsets in Foundations of SQL, ensuring the analysis is efficient and only ingesting the required data.



## **Module 10: Data Science for Policy**

This module takes a boot camp approach to taking a project through from scoping to analysis. This course uses the R framework to help embed prior learning in agile project management, data ethics, data visualisation and effective communication in the form of a written report. The focus is on communication of technical content appropriate to the target audience – policy makers.

## **Year two**

Organised events, workshops and seminars stimulate and encourage further learning. You will work on specific projects that utilise your learning from the first year.





## Eligibility

We look for candidates who have a passion to work in Data Science, and who have some programming ability in R, Python or another data analysis relevant language.

You must have, or be expected to get at least one of the following:

- a 2.1 degree, or post graduate qualification, in a discipline with numerical or statistical elements.
- equivalent wider experience.

The **Graduate Programme** is an exciting opportunity for you to gain **valuable data science skills**, and experience in applying these to **real world problems**.







I came to the scheme with a degree in Geography (Geocomputation and Spatial Analysis) with geospatial analysis skills in Python and experience in using GIS software.

As a naturally curious person, the **sheer breadth of projects** within the Civil Service excited me as I knew I would be able to interact with different data science domains. As someone beginning their career, the graduate programme's focus on learning and personal development also attracted me.

I am currently working on a COVID-19 project exploring a hotspots risk model. I have also volunteered to analyse case studies for the Data Science Campus faculty to provide input into the machine learning course they are developing.

I have become a more **confident programmer** through the graduate programme and I have learnt a lot from the programme lecturers and the people I've worked with. This has helped me to develop my data science skills in new domains, such as machine learning, at pace. I will continue to develop my programming and professional skills over the remainder of the programme and it is a **great investment in my future data science career**.



**Arif Ali**

Graduate Data Scientist



The graduate programme appealed to me because I wanted the opportunity to experience a variety of different data science projects within the public sector.

I had completed an Undergraduate degree (BSc) in Mathematics at Durham University, and a Masters (MSc) in Data Science and Analytics at Cardiff University and came in to the programme with a good foundation in statistics and data visualisation, as well as basic Python and R programming skills.

Throughout the programme I have **gained confidence** and built on the practical skills needed for a data science project. I have worked on building the Office for National Statistics' (ONS) capability in distributed data processing and analysis and used text analysis to automate occupation and industry classification for the Census.

**I also had a fantastic opportunity to go on a secondment to 10 Downing Street to work on visualising and communicating data and statistics for senior politicians and advisors.**

I have now been **promoted** to a Data Scientist role in the Government Statistical Service's Good Practice Team.



**Rowan Hems**

Graduate Data Scientist

Our Graduate Programme vacancies are advertised via:  
[www.civilservicejobs.service.gov.uk](http://www.civilservicejobs.service.gov.uk) and on social media:

 **Office for National Statistics**  **[@ONSrecruitment](https://twitter.com/ONSrecruitment)**

To find out more information on what we do at the  
Data Science Campus visit our website:

[datasciencecampus.ons.gov.uk](http://datasciencecampus.ons.gov.uk)

