Data Science Campus

Data Science Graduate Programme

2022 to 2023

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

What will the Graduate Programme offer?

The digital revolution is allowing public sector bodies to use innovative and experimental data sources to provide new insights. The Data Science Graduate Programme looks for talented individuals with a passion for data, to help us shape the future of data science in government and the wider public sector.

Now in its third year, with graduates in 10 different public sector organisations, the programme provides successful applicants with highly valuable work experience, that develops their knowledge and skills through our structured two-year learning pathway. Graduates join experienced teams of data scientists in UK government or public sector organisations. In these teams the graduates will use their training gained during the two years, to apply cutting-edge data science techniques to assist policy makers.

The two-year Graduate Programme has been designed to include:

- Training in core data science skills
- Project based learning
- Professional development support
- Opportunity to engage and learn with peers from different organisations
- Attending data science conferences and seminars

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 online, working through independent learning or reading material, and/or exercises. Between these learning sessions **graduates will be working in their home organisation** applying their learning on projects.

The first year of the programme allows the graduates to spend structured time thinking about and practising data science. They will establish a **core set of data science skills** that will raise their capability to contribute to their home organisations. They will have the opportunity to **build collaborative relationships** with other graduates and leading UK government Data Scientists.

During the second year of the programme the graduates will work on **projects that utilise their learning from the first year**. They will attend a series of events such as hackathons, that encourage their Data Science skill development.

At the end of the programme, the graduates will have **acquired new, highly sought-after analytical skills**, forming a secure foundation for a future career in public sector data science. They 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 to nurture their passion for cutting edge data science techniques**, equipping them for a job with purpose and a stimulating career.

Curriculum Overview

The Graduate Programme curriculum provides a robust 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 graduates with a range of tools and experience in applying them, ensuring that they 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 2022 to 2023.

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 graduates are expected to complement their programming skills by choosing the framework in which they 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. Graduates 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 programmers in developing robust, well-integrated programs that achieve maximum benefit from programming software. Modular code helps Python and R programmers to consider the utility and structure of their 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, covering the theory and applications of widely applied machine learning techniques. Programmatic implementation of supervised and unsupervised learning techniques is then taught, using Python or R programming frameworks.

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. Participants will apply the knowledge taught in the curriculum and self learning in a range of fields including machine learning with satellite data, network analysis of economic industries and sentiment analysis of government related social media data.

Module nine: Databases & Big Data
This module's aim is to equip graduates with the skills required to apply their prior learning to big data. This includes querying databases to return the required data subsets in SQL, ensuring the analysis is efficient and only ingesting the required data. Participants will get experience with distributed computing using the PySpark or SparklyR frameworks.

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 encourage further learning. Graduates work on specific projects that utilise their learning from the first year.

Impacts and benefits of the programme

Graduates

- Provides them with the learning opportunities to think critically about data and data science, with appropriate consideration of ethics, impartial analysis and effective communication of data-derived insights.
- Enables them to produce work of a high standard which has a real world impact.
- Prepares them for a successful data science career in the public sector.

Home organisation

- Increased data science capability through highly motivated and enthusiastic graduates.
- Promotion and retention of skilled data scientists in the public sector.
- Increased collaboration across the public sector in resolving real world problems.

Eligibility

We look for candidates who have a passion to work in the data science world and have existing knowledge and experience of data science techniques.

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

- a 2.1 degree in a relevant discipline
- post graduate qualification in a relevant discipline

  OR
- equivalent experience

The Graduate Programme is an exciting opportunity for graduates 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 worked on a COVID-19 project exploring a hotspots risk model and also analysed case studies for the Data Science Campus faculty to provide input into the machine learning course they developed.

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. The programme has been a great investment in my future data science career.

Arif Ali
Graduate Data Scientist (2021 cohort)

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.

The programme gave me confidence and built on the practical skills needed for a data science project. I helped build 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 was also promoted to a Data Scientist role in the Government Statistical Service's Good Practice Team.

Rowan Hemsi
Graduate Data Scientist (2020 cohort)

Our Graduate Programme vacancies are advertised via: **www.civilservicejobs.service.gov.uk** and on social media: **Office for National Statistics** **@ONSrecruitment**

For queries about bidding for places on the Graduate Programme contact: **datacampus@ons.gov.uk**

To find out more information on what we do at the campus visit our website: **datasciencecampus.ons.gov.uk**