

Special feature: How GAD uses data science to enhance our service offering

Data science is the combination of programming code and statistical knowledge to extract understanding and insight from data.

Data has always been a key part of the work of actuaries, and harnessing data science techniques enables GAD to work more efficiently and maximise the value of data for departments across government.

For example, the use of data science techniques allows us to use [reproducible analytical pipelines](#) to handle repeatable tasks effectively, and the use of modern coding languages allows us to create 'self service' dashboards for clients.

GAD recognises the important part that data science can play in how we process, analyse and gain insights into the high volume of data used in actuarial work.

As part of the [GAD 2025 Strategy](#), we have committed to investing in data science expertise for our actuaries and analysts to ensure that GAD's clients the highest quality of advice possible. In the rest of this article, we set out how our clients have already begun to experience the benefits of this commitment.

Pensions dashboards

Building on the increased programming expertise required for data science, we have developed several dashboards which are used internally to automate processes which are repeated regularly. We have 3 main dashboards that we use for the public sector schemes that we manage.

We have created the dashboards using the open-source programming language [Python](#). The dashboards use centralised code, so any changes to the model only need to be made once (rather than for each individual scheme). This saves a significant amount of time and ensures that our approach to all schemes is consistent.

1. Valuation calculations dashboard

By building these calculation routines into a structured model, we are able to ensure consistency of quality across all of the public service pension valuation calculations, rapidly update results for changing assumptions and instantly generate visual representations of the actuarial calculations.

This is brought together in our easy to use [valuation calculations dashboard](#). It consists of drop-down menus and inputs which allow the user to select

which scheme they wish to work on and what kind of analysis they want to carry out.

2. Analysis of experience dashboard

The [analysis of experience dashboard](#) is used to analyse changes in membership between valuations. It looks at how many members have exited the scheme and how that compares with what was assumed. This analysis allows us to adjust our assumptions going forward to reflect changing membership characteristics and ensure that appropriate contribution levels remain in place.

3. Retirement calculator

The McCloud legal ruling gives some pension scheme members the choice between benefits in two different types of pension scheme.

GAD's [retirement calculator](#) has recently been upgraded by our data science specialists to improve the user experience. It uses Python to project the benefits that people will get at retirement in both schemes. This allows members to view what their own personal benefits will be and make an informed decision on which benefits they choose.

National Situation Centre

In addition to our in-house dashboards that have been developed, we have seconded staff to the National Situation Centre (SitCen). SitCen predict and manage developing crises in the UK. Our staff have used actuarial and data science skills to develop dashboards which support their analysis. Read more about the secondment experience of one GAD member of staff in [a blog by Sean Laird](#).

Heat decarbonisation plans

The UK government has committed to reaching net zero carbon emissions by 2050 or earlier if possible. Heat decarbonisation involves reducing the amount of carbon produced by central heating systems. Given the cold weather that we experience in winter, reducing the carbon footprint of our heating systems will be key in our move towards net zero.

The Department for Education (DfE) commissioned external consultants to produce plans for 205 schools to help with the [decarbonisation of school estates](#). The heat decarbonisation plans received for each school contained a variety of Word, PDF and Excel files.

DfE asked GAD to help support its work by:

- creating a database to collate information contained in the heat decarbonisation plans
- extracting the relevant information from the set of documents
- designing a set of criteria to identify which schools to target
- assessing the project against the government's levelling-up policies

GAD worked closely with DfE to agree the scope of our work and the timeline for delivery of each project stage.

We met several issues when extracting and processing the data from the heat decarbonisation plans provided, such as:

- report formats which varied by school
- documents that were not searchable
- different names being used for the same school across different documents
- figures not always reconciling across documents

To address some of these issues, we used data science techniques to extract and manipulate data.

A 'fuzzy matching' technique was used to map school names that were not identical but were likely to refer to the same school.

Fuzzy matching: A technique for determining how similar 2 character-strings are to each other. It is usually based on the number of multi-letter chunks that they share. For example, 'washing' is highly similar to 'wasting' because the two words share the chunks 'was' and 'ing'.

It is often used to link datasets where identifying labels may have minor spelling mistakes, so relying on a perfect match would lead to errors.

Our team also used a 'data scraping' algorithm to automatically cycle through the documents received to extract specific information.

Data scraping: Extracts information from a document or web site and exports it into a spreadsheet or other file. For example, wording from PDF documents can be automatically placed into a spreadsheet so that it can be manipulated and analysed. It can also be used to extract information held on web sites for investigation.

Not all the required information was available at the start of the project. Having a process that was robust and easy to reproduce and iterate was crucial. We were able to re-run the exercise quickly when more information came to light.

None of this would have been possible without the expertise of GAD's data scientists.

We provided the client with a clean, centralised database of key data fields from the heat decarbonisation plans. We also produced a ranking of all schools based on decarbonisation priorities. This information allowed DfE to make efficient decisions on funding allocation to schools.

Collaboration with universities

GAD has been working with UK universities this year to suggest data science projects being carried out by students. The projects contribute towards the students' Masters degrees.

Each project challenged the student to use data science techniques to analyse real world data. Students use techniques such as linear regression and machine learning to analyse the data and derive insights. Actuarial staff from GAD provide oversight and feedback on their work throughout.

The University of Manchester, City University and University College London have taken part this year. The projects we proposed were based around public service pension scheme data collected by GAD during recent projects in this sector.

Unprocessed data from the 2016 and 2020 valuations of the cost of providing public sector pensions is used by students as the basis of some of the analysis.

Areas of focus included:

- Geographic analysis – do the characteristics of schemes differ between different UK regions? For example, do salaries and member behaviours vary in practice between regions? What factors might explain any patterns found?
- Workforce changes – how do members progress between different job roles or decide to leave over time? Can we predict the demographics of the scheme in 2024 based on this?

The results of this analysis have been used to support workforce planning and recruitment.

Our [relationships with universities](#) are ongoing and we continue to support the development of data science experts.

Moving forward

Data science is a growing area of GAD's expertise. Whether it is [Python](#), [R](#), [R Shiny](#) or [JavaScript](#) – the list of programming languages is a long one. At the same time we are deploying advanced statistical methods to improve our understanding of the issues faced by our clients, enabling more impactful advice, and ultimately better outcomes for the UK.