Reproducible Analysis in the UK Government

A key function of government statisticians is to produce official statistics for publication. Often these statistics have a direct impact on government policy, so it is imperative that they are accurate, timely, and importantly: reproducible. At any point in the future, we should be able to reproduce all the steps required to produce a statistic, but manual processes (common in many official publications) can make this challenging. This presentation is about a new approach to producing official statistics using the R language that was trialled in the UK Government, and has become a major success story.

Taking inspiration from the fields of reproducible research and DevOps, an approach called 'Reproducible Analytical Pipelines' (RAP) was trialed in the Department for Digital, Culture, Media, and Sport (a central UK Government department). The proof of concept involved taking an existing, highly manual, official publication and creating a reproducible pipeline that allowed the publication to be reproduced instantly. A clear audit trail and high standard of quality assurance was maintained by using version control, test driven development, and continuous integration. Finally the pipeline (comprising a bespoke R package and Rmarkdown document) was published as open source software, allowing complete transparency in how the statistics were produced.

We publicisied this proof of concept with a blog post, and were able to follow up with a second blog post as the project was adopted in other government departments a few months later. We also produced an ebook (written in bookdown), a website, and a massively open online course to help spread the message in the UK Government and elsewhere. Within a year, the project had been endorsed by the UK Statistics Authority, the body responsible for assessing the accuracy of official statistics. As of April 2019, the project has been so successful, that there are now 22 RAP projects under way and 42 champions across 24 UK Government bodies. We even have a hex sticker.

In this presentation we will share some of the technical details behind Reproducible Analytical Pipelines, and how we built a community of champions to take the work forward. We hope that this presentation will serve as an example of how innovation can be fostered within a very large and traditionally conservative organisation, and how modern tools and practices can be used to reduce repetitive and onerous manual labour, improve accuracy, and reduce the time taken to produce official statistics.