



CASE STUDIES SERIES #001



Open Source
Data Analysis
for Councils

Alistair Herbert / Data to Insight

James Rapkin / Southend-on-Sea Borough Council

May 2021

In 2018, data teams at Southend-on-Sea Borough Council saw the potential of **open source analysis tools like R** to enhance the council's data work. Three years later, that vision is **generating daily value** for services across children's and public health departments. How did it happen, and what good is it doing?

Why open source?

The team needed **better approaches to existing problems**, and they knew those approaches existed. With staff in place who knew how to use R, this was the obvious place to look first, and it generated real benefits.

The fact that the tools were **free to start learning**, and affordable to license during widespread implementation, made procurement easy to justify on a cost basis, and made it possible to generate real examples of value before purchasing anything. **Final costs were low** compared to traditional competitors.

Crucially, the open source tools were just the **tools which did what the team needed to do**. Open source software is no longer a niche enterprise – multinational companies rely on R, Python and similar languages, and there exist proven and popular code distributions which they can trust in their corporate environments.

The learning process

The team **already understood the value of exploring new possibilities** – they spent previous years working with Excel, VBA and SQL to automate data handling where possible, and had already built positive relationships with the services managing their databases. When they had the chance to recruit a new colleague, they saw **existing knowledge of R** as a huge opportunity, and pursued skills transfer approaches to help that knowledge spread within the team. Preparation met opportunity.

Skills transfer within the team was therefore the key learning method, with informal training delivered as part of a regular schedule of "Friday morning" development sessions. Colleagues took away the basics, then learned by trial, error, and plenty of questions. Some team members also subscribed to online learning platforms. In all cases, though, the primary focus has been on **learning by doing**, and on protecting the time needed for colleagues to pursue their learning and turn it into real value for the organisation.

R wasn't for everyone – as with any tool or method, a subset of colleagues found it useful for their work. The team embraced this: different people doing different work often need different tools to get the job done.

The current mix

R and Shiny complement the existing reporting environment in Children's Services. Internal day-to-day reporting remains in SSRS – a secure, stable environment which all front-end users already understand. But R is becoming an **essential tool backstage** for data processing and integration – especially with external sources.

R is also used for **bespoke analysis projects** and connected data work; for one example, reporting for the local safeguarding partnership, where multiple partner organisations provide data which the team processes using R scripts, and presents via a Shiny app which everyone can easily use. Projects like this have been extremely popular, both with users and developers, and have driven more interest in further use of the R/Shiny toolset.

Colleagues share scripts and perform quality checks on each other's work. A code library is slowly emerging from this work, with scripts stored in a central accessible location. The process has increased levels of discussion between analysts – "has anyone else attempted..." is a common question which often generates shared learning opportunities.

The team are now skilled enough that they can support each other's apps, digging into existing code to make changes where necessary. Reasonably clear coding standards exist within the team, which aids legibility, future development, and quality checks. They are also working on lots of **new ideas** – they're putting together an interactive benchmarking tool which draws in published data for easy analysis, and they're **beginning work in Python** as their next open source journey. Python integration is supported by Esri ArcGIS, offering new possibilities for working with geographic data

What good has R done?

Children's Services leaders are not only getting faster information, they're getting better insights. Automations which free up analyst time mean that the analyst can spend less time wrangling with data and more time interrogating it and looking for the stories data is telling. These stories are why leaders come to data in the first place, and they lead to better decisions. Ultimately, children receiving services are the beneficiaries of this innovation.

The **COVID-19 response team** saw the benefits of this in tangible ways, as R automations were used to turn around presentation and analysis requests on a daily basis, generating real-time insights which were crucial to the local response work. In **Public Health**, it's now common practice for analysts to clean public datasets using R, analyse the resulting data using R, and output R visualisations into report documents as a matter of course. And this is within 2 years of that team first encountering the tools.

Then there are the **organisation-wide benefits**. A Shiny app designed to process postcode updates removed longstanding issues with redundant spreadsheets. Where previously numerous postcode files in different areas were updated separately, the new Shiny process ensures that all postcode-related work in the council is accurate and timely. And with automated processes, if mistakes do occur, they can be traced in code and corrected – a luxury not afforded by manual processing.

Working with IT

The first step happened without IT involvement; one colleague installed R locally and created some **exemplar work** to show off. Then they went to IT to ask for a corporate install. It's fair to say that the IT service were skeptical at first. But, when they saw the kinds of things that the team were capable of doing with R, as well as a whole range of examples of **other government agencies** which were already using Python and R, they took the request seriously and found a way to make it happen.

It took a little while to get the installation right – there was some initial confusion over which software was requested, and some issues with firewalls blocking publication of Shiny apps. But these were resolved, giving the whole team access to R Studio.

The next step was to use Shiny to publish visualisation and analysis apps based on the R data. This was a case of **demonstrating capabilities to service leaders**, who saw the potential benefits and said yes. The team then worked with IT and procurement partners to establish the necessary licensing and infrastructure changes

Any regrets?

The team want to improve their version control. They also would have benefited from establishing a shared code library earlier, capturing more of their early work within it. But there's also a feeling that this might have been a necessary part of the journey: changes like this can't always happen overnight, and it may have been necessary to adopt individual elements of their open source strategy – beginning, of course, with the core language of R – rather than implement wholesale change.

If introducing R had required upfront commitments around code libraries, version control, QA, coding standards, and learning programmes, even before the first beneficial products emerged, would people have come on the journey so enthusiastically? It may be that they needed to demonstrate value in small ways before approaching these broader process changes.

Telling the story to stakeholders

There were two crunch points in Southend's journey: first, persuading IT colleagues to whitelist the relevant tools, and find time to install them. Second, persuading budget holders to purchase Shiny as a visualisation and interaction companion to R code.

In both cases the story was about being able to do things quicker, and about knowing the data was right. Whether integrating datasets from multiple safeguarding partners into a single analysis app, or consolidating departmental postcode data into a verified single dataset, the team's work demonstrated to stakeholders that this way of working *just worked better* than their previous approaches. The data was more reliable, the analysis was more timely, and the processes were auditable. The perceived risks were more than manageable – in many cases, misperceptions came from the unfamiliarity of the new.

For more information, please contact:

datatoinsight.enquiries@gmail.com

www.datatoinsight.org

twitter.com/datatoinsightUK