

Providing Software Support to Enable Research: From Feral Parakeets to the Times Digital Archive

Stephanie E.M. Thompson
IT Services, University of Birmingham

Simon J. Branford
IT Services, University of Birmingham

Richard D.J. Bufton
School of Biosciences, University of Birmingham

Aslam K. Ghumra
IT Services, University of Birmingham

Laurence A. Hurst
IT Services, University of Birmingham

Simon J. Thompson
IT Services, University of Birmingham

Andrew C. Edmondson
IT Services, University of Birmingham

Abstract

With the increasing interest in Data Science and demand from researchers for research software support, the University of Birmingham has recently established a Research Software Group (RSG) as part of Advanced Research Computing (ARC). Whilst the Research Engagement Group provide advice on storage and research data management, the RSG provides more advanced help similar to that provided by Data Stewards in Dutch research institutes, to build project-specific databases and design data workflows. Some of the Research Software Engineers (RSEs) are funded by individual schools or Colleges, which allows them to specialise in that research area and provide dedicated support.

The range of services provided by ARC are branded as BEAR (Birmingham Environment for Academic Research). There are currently eight members of the RSG whose aim is to provide specialist software support. Rather than create the finished product for a researcher, the group prefers to coach them so that the researcher is upskilled and can then complete similar projects in the future. Training workshops are provided in Software and Data Carpentries as well as NVIDIA Deep Learning. They

Submitted 05 July 2019

Correspondence should be addressed to Stephanie Thompson, IT Services, Elms Road Data Centre, University of Birmingham, Edgbaston, Birmingham, B15 2TT. Email: s.e.m.thompson@bham.ac.uk

The 15th International Digital Curation Conference takes place on 17-20 February 2020 in Dublin, Ireland

URL: <http://www.dcc.ac.uk/events/idcc20>

Copyright rests with the authors. This work is released under a Creative Commons Attribution 4.0 International Licence. For details please see <http://creativecommons.org/licenses/by/4.0/>



also provide High Performance Computing (HPC) support for our cluster (BlueBEAR) and for using BEAR GitLab, which allows version control and tracking for software projects as well as providing an easy way to share code.

One such coaching project, which was recently featured on BBC Springwatch, was the development of a website and mobile application allowing members of the public to record sightings of feral parakeets in the UK and around the world. The PhD researcher (Richard Bufton), had no previous experience of coding so our RSE started by teaching Python and fundamental programming principles. They then designed a database to store the data entries and developed the web-based infrastructure to support a mobile application (see <https://parakeetsightings.bham.ac.uk/>).

For researchers who have some experience in coding, the RSG can provide advice. The Times Digital Archive (TDA) contains an abundance of data that can be analysed, however the TDA only provides a scanned article of varying quality. The researcher had access to a copy of the full text of the TDA and with the support of an RSE, has been processing it to remove unnecessary XML tags. To investigate how surveillance is portrayed in texts in history, thousands of text files needed to be processed containing approximately 1.5 billion words. If the researcher had used the original setup it would have taken forty days to process, after an RSE helped optimise the jobs on BlueBEAR it ran in forty hours.

The RSG has rapidly grown over the last two years to provide a service which is in high demand. In May 2019, there were 7 advice sessions, 23 coaching engagements, 10 coding engagements and 100 new tools/applications were installed on BlueBEAR. Going forward, the University has recently purchased the largest POWER9 Artificial Intelligence (AI) cluster in the UK and support is being provided for pilot projects, including porting code to the new architecture.