Newcastle University

Research Software Engineering

ANNUAL REVIEW 2022

Introduction

2022 will be remembered as the year we hosted the RSE community in Newcastle for RSECon22 – a week of incredible talks, workshops and networking events. I also see it as a transition year; driven by both staff turnover and a need to have more control over our project and financial processes. As expected, we've continued our pattern of growth via recruitment, though this has been balanced by staff leaving for new opportunities. Of particular importance is the recruitment of our first Operations Administrator – a role which will transform the way we organise ourselves and engage with the wider University.

Looking ahead to 2023, we want to continue our progress towards formalising some of our internal processes. What was once a nimble start-up mentality that served us well, has begun to cause issues that we could avoid with carefully applied processes, taking care not to lose our flexible approach to the work. We also want to look for areas of the University where we can do more to support research and become part of "business as usual" for any computational research done at Newcastle. Currently opportunities around HPC and training are the most obvious areas to explore, but there will be others. In 2022, the team had a project income of £2.04m (up 29% from 2021) from 50 different projects, with a total research value of £67.19m to the University.

Mark Turner
Head of Research Software Engineering

Contents

2022 Review

Highlights

Monthly highlights and key events from the first half of the year

Jan

Jannetta Steyn is awarded a Software Sustainability Institute Fellowship.

Feb

Mark completes his term representing Newcastle on the N8 CIR Steering committee.

Mar

Team restructure is completed, Becky and Kate are promoted into senior roles.

Apr

Robin Nandi and David Halliday join the team.

May

The team attends the Thinking Digital technology conference at Sage Gateshead.

Jun

Iolo Squires joins the team on a PhD placement and works with Dave Horsfall in the Hannifa Lab.

Highlights

Monthly highlights and key events from the second half of the year

Jul

Kathryn Garside, Tiago Sousa Garcia, Becky Osselton & Robin Wardle nominated for NCL Success Awards.

Aug

Michelle Gilbride joins the team as our first Operations Administrator.

Sep

Newcastle hosts 300 delegates for the UK Research Software Engineering conference.

Oct

Richard Howey and Bowen Li join the team. Robin Nandi elected to the RSE Society board of trustees.

Nov

Imre Draskovits joins the team. The team starts a University-wide speaking tour.

Dec

Mike Simpson, Jannetta Steyn & Becky Osselton join the RSECon23 organizing committee.

RSECon22

RSE

Between September 5th-9th we were honoured to host the annual RSE conference at Newcastle University. This was a return to an in-person conference after the pandemic so we built a varied technical programme around several networking events to allow the community to reconnect. We were delighted with how well it turned out and the feedback from the community has been excellent.

ORACLE for Research · Red Hat · epcc · Software Sustainability Institute · The Alan Turing Institute

Finances

Facility Funding Model



Direct Costs + Estates + Non-Salary

Team Size x 220

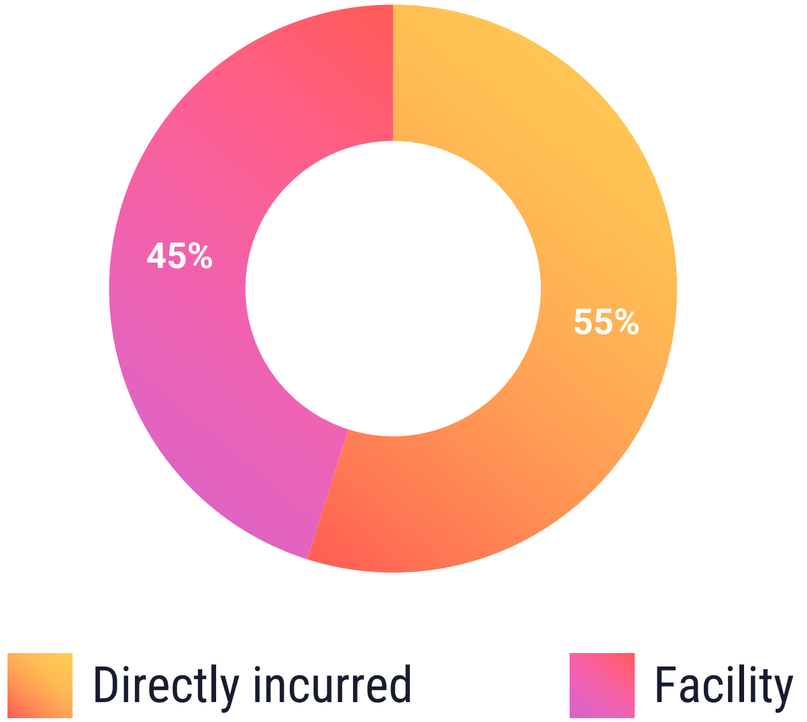
=

Day Rate

The final figure represents the cost of running the team in a way that breaks even after 85% of annual capacity is charged out.

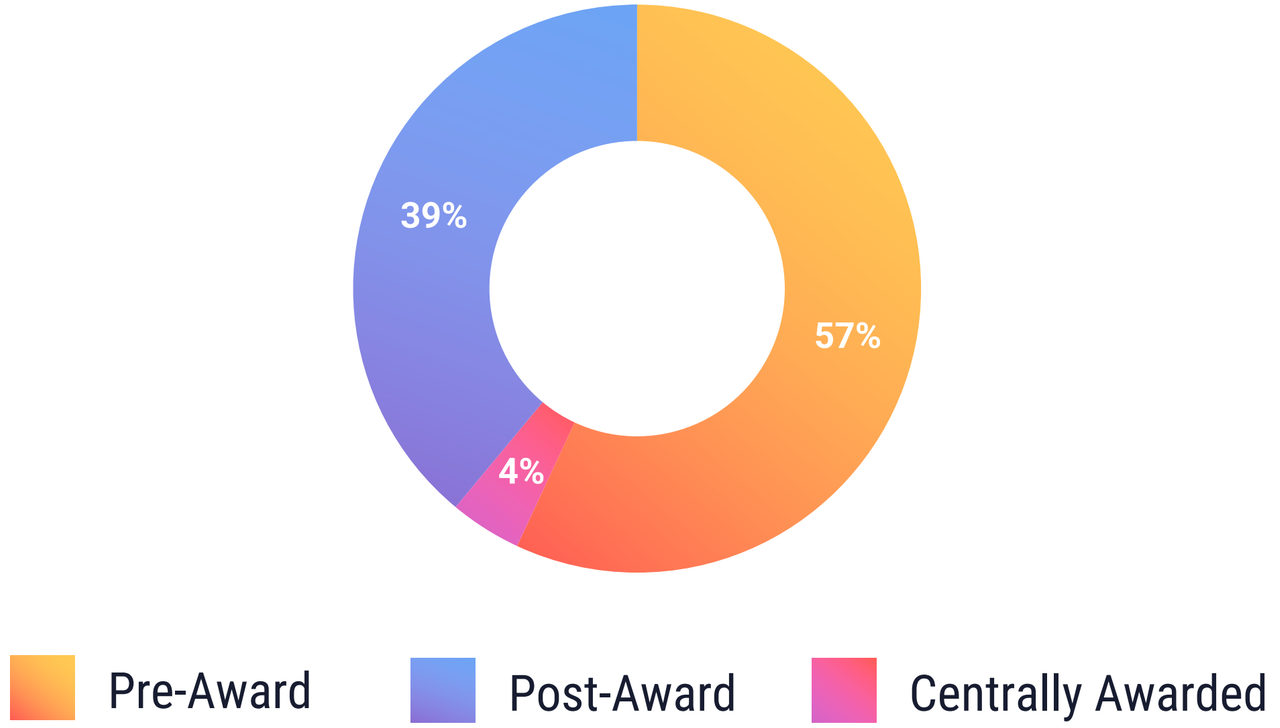
Income by Funding Model

Income and the number of projects using the facility rate costed under the directly incurred funding model have been close to equal this year. This represents progress towards using the day rate for costings as previously costed projects start to be completed. Our income from projects costed using the day rate has increased by 36% since 2021, which is good progress towards our goal for the majority of our funding coming from this source.



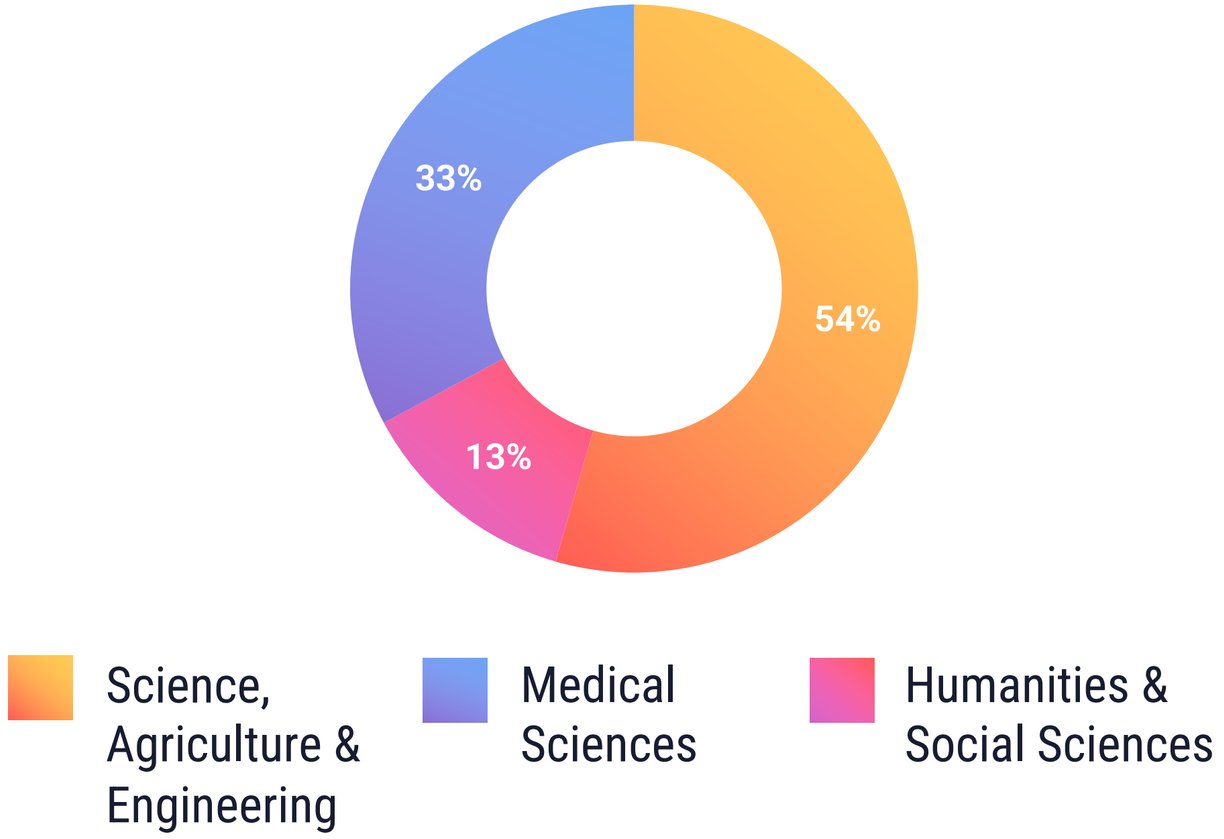
Income by Award Stage

Since 2019 there has been a year on year increase in the percentage of our income that comes from projects that have already been awarded funding (‘post-award’). This is likely in part due to researchers who have existing grants being made aware of the team from colleagues. However, it is probable that a reasonable amount of post-award work is still a consequence of Covid impacting on researcher’s plans and travel budgets. A small number of projects have been centrally awarded this year as RSEs contribute to teaching and internal initiatives.



Income by Faculty

Income from FMS has increased in value from 2021 in percentage terms. Income from HaSS has decreased in percentage terms again this year. In terms of the number of projects there was only one additional project from FMS compared to HaSS indicating they tend to be of greater value per project. We undertook more projects from SAgE than the other two faculties combined, and income in percentage terms was still similarly the largest for the team.



The Team

Kate Court

Senior Research Software Engineer

Head of Middleware & Data Science

Imre Draskovits

Research Software Engineer

Web & Mobile

Tiago Sousa Garcia

Research Software Engineer

Web & Mobile

Kathryn Garside

Research Software Engineer

Data Science

Michelle Gilbride

Operations Administrator

Management

David Halliday

Research Software Engineer

Web & Mobile

David Herbert

Research Software Engineer

Web & Mobile

Dave Horsfall

Senior Research Software Engineer

Web & Mobile

Richard Howey

Research Software Engineer

Middleware & Data Science

Bowen Li

Research Software Engineer

Middleware & Data Science

Daniela Basurto Lozada

Research Software Engineer

Web & Mobile

Robin Nandi

Research Software Engineer

Middleware & Data Science

Nik Khadijah Nik Aznan

Research Software Engineer

Middleware & Data Science

Becky Osselton

Senior Research Software Engineer

Head of Web & Mobile

Mike Simpson

Research Software Engineer

Middleware & Data Science

Jannetta Steyn

Senior Research Software Engineer

Head of Training Delivery

Frances Turner

Research Software Engineer

Middleware & Data Science

Mark Turner

Head of Research Software Engineering

Management

Robin Wardle

Research Software Engineer

Middleware & Data Science

Projects

NHS DIGITAL

ePrescribing Risk & Safety Evaluation Tool

Stephanie Klein

Newcastle upon Tyne Hospitals
NHS Foundation Trust

Challenge

ePRaSE is a self-reporting tool available to pharmacists within the NHS to evaluate the ePrescribing system used in their trust. Such systems attempt to codify rules for drugs and should intervene when a medical practitioner attempts to prescribe combinations of drugs that could lead to negative health outcomes for the patient.

Outcomes

The ePRaSE tool evaluates system responses and assesses the potential risk to test patients. It captures and compares data from all trusts to provide an overview of the state of ePrescribing in NHS England.

MEDICAL RESEARCH COUNCIL

VOICE Platform

Richard Airey

National Innovation Centre for Aging

Challenge

VOICE is based in the UK's National Innovation Centre for Ageing (NICA), a world leading organisation created with a £40 million investment from UK Government & Newcastle University.

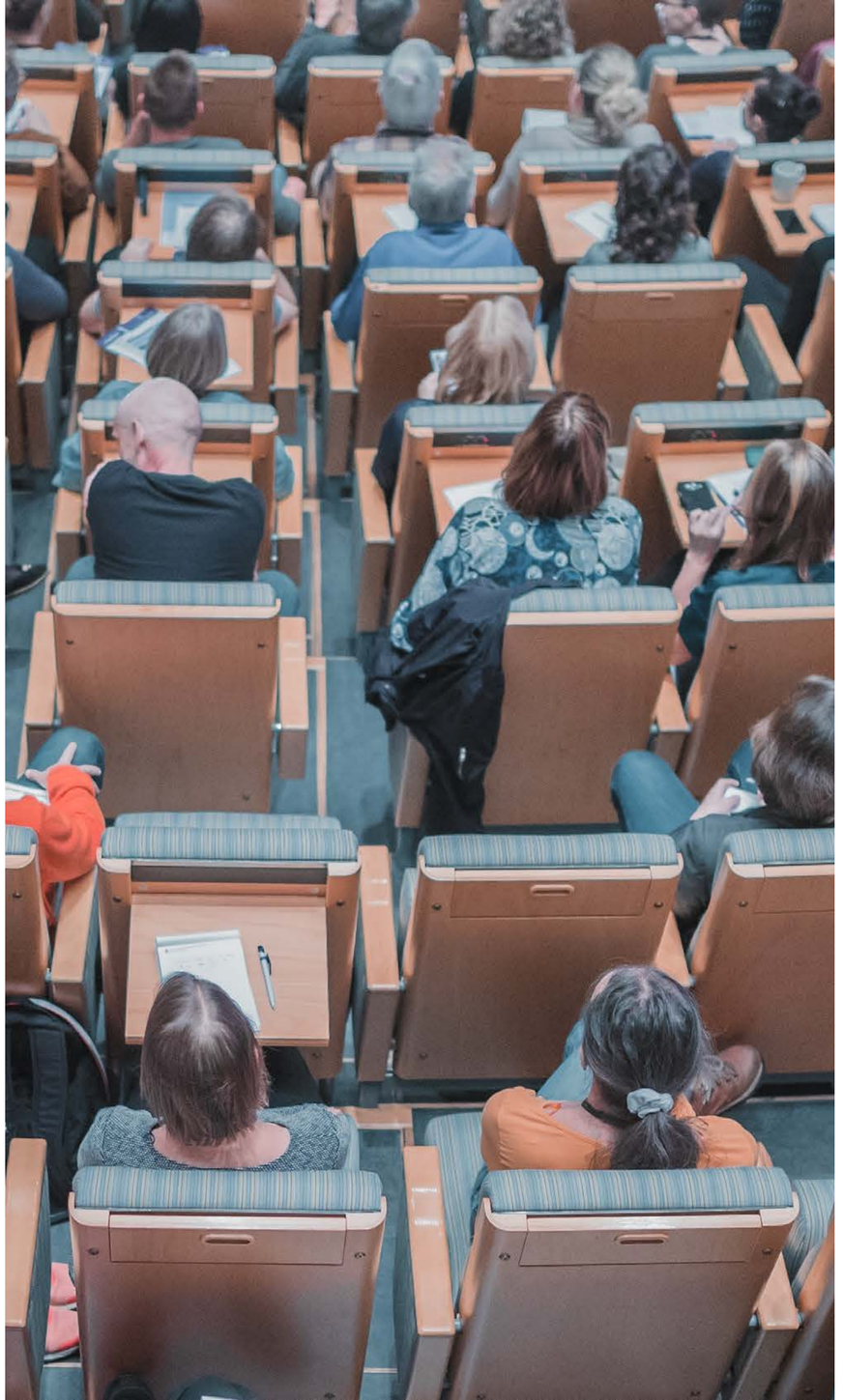
Outcomes

NICA and VOICE are leading dialogue with citizens on what is needed for healthy ageing. They work with business partners to co-design and develop market desirable products and services intended to transform lives.

OFFICE FOR STUDENTS

Institute of Coding

Steve Riddle
School of Computing



Challenge

The team provide teaching support on a key module that forms a part of a technical apprenticeship programme for the Institute Of Coding.

Outcomes

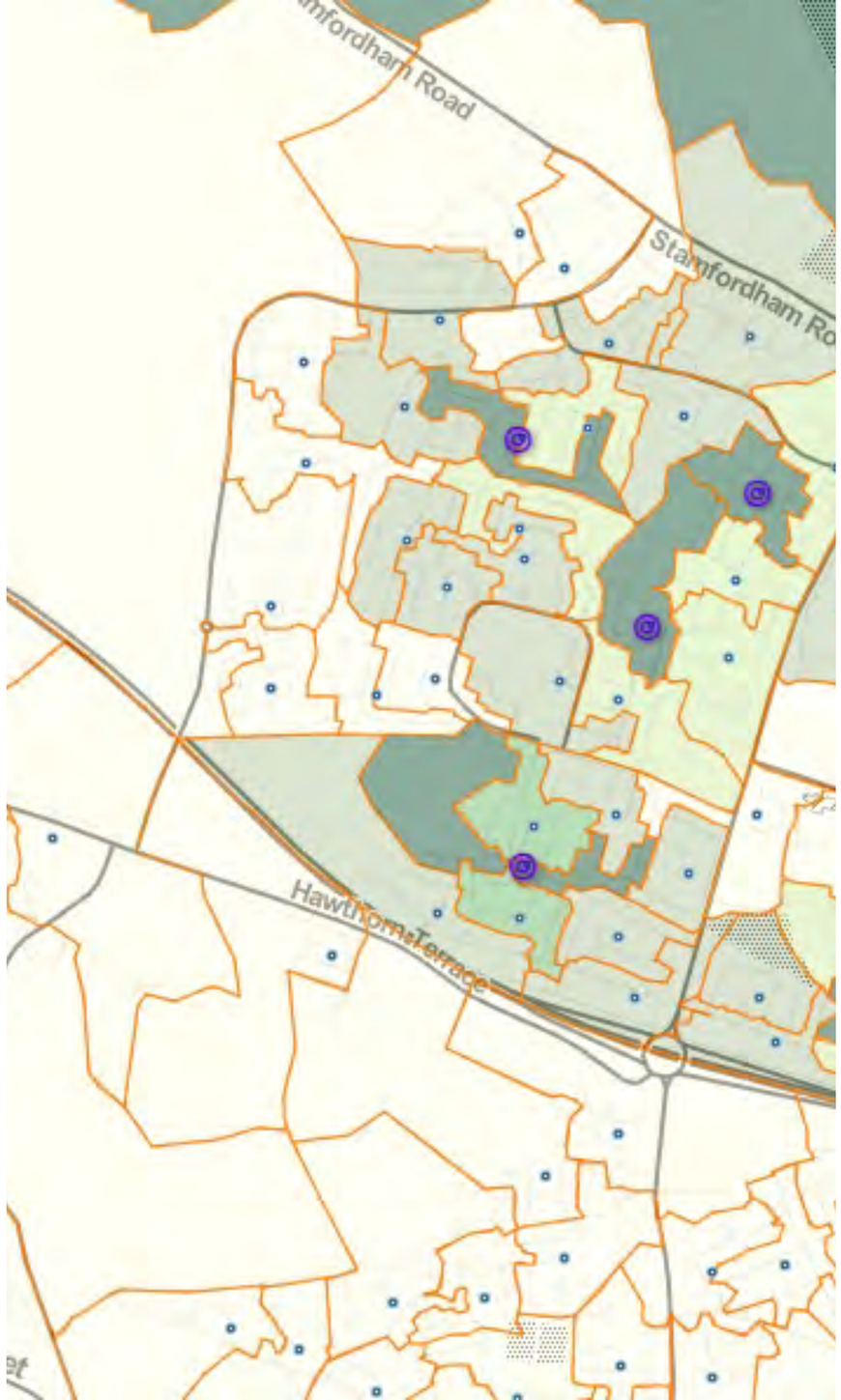
Teaching is delivered in a blended format, mixing online and video content in Canvas with hands-on practical coding activities in workshop sessions.

THE ALAN TURING INSTITUTE

Spatial Inequality and the Smart City

Rachel Franklin

Centre for Urban and Regional Development Studies



Challenge

Our cities are now smarter than ever, but they are also more unequal. Where sensors, such as those measuring air quality, are placed can either help tackle or contribute to this inequality. When decision makers place sensors there are a number of hazards, risks and vulnerabilities in the environment and local population to consider.

Outcomes

This project resulted in a decision support tool that presents population data and suggested sensor networks to the user. A genetic algorithm is used to evaluate the coverage of networks for different population groups. The tool is designed as a thinking tool to help visualise the gains and losses of networks, highlighting the complexity of reaching a ‘fair’ smart city.

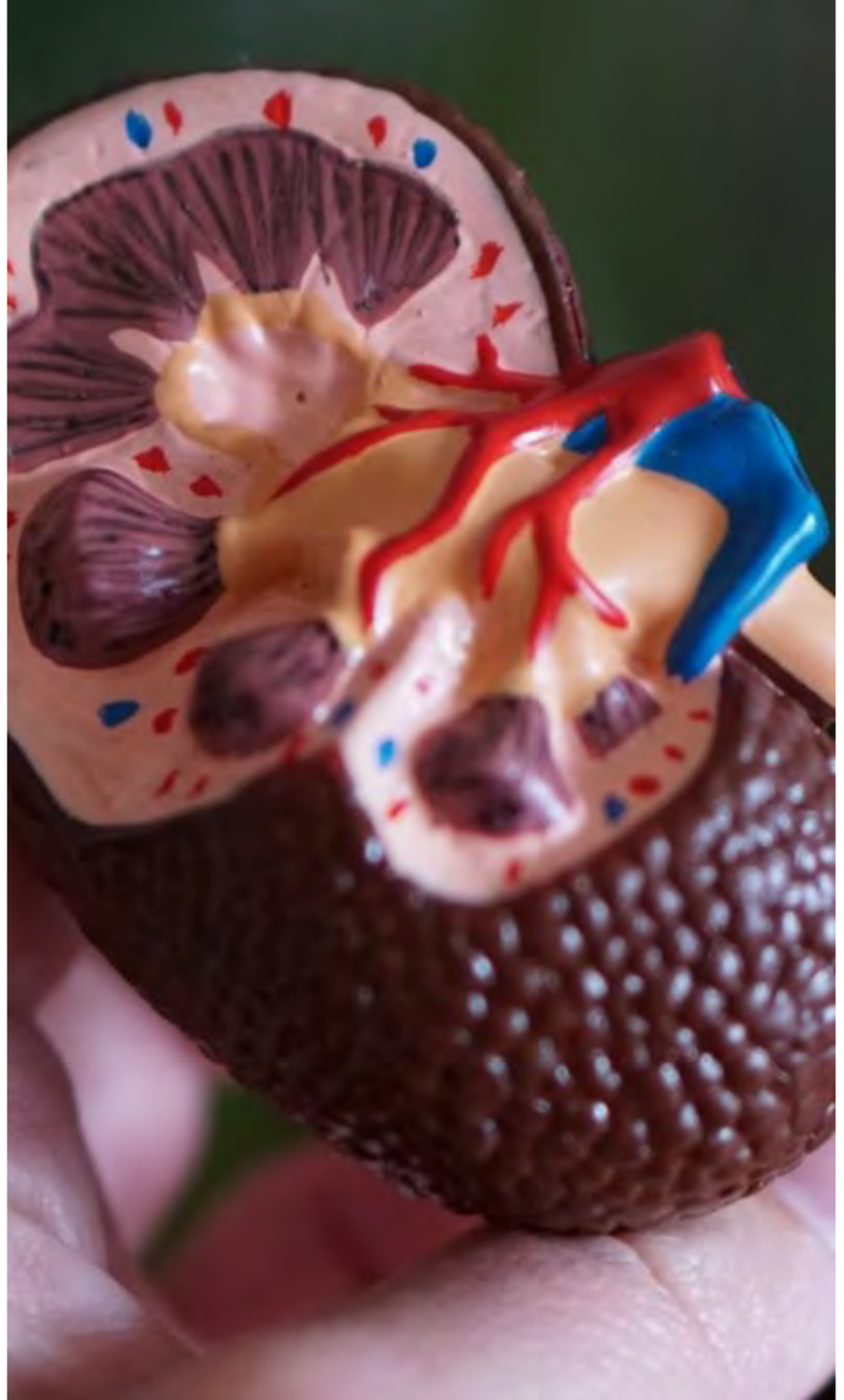
Court, K. & Roberts, J. (2022, Sept 6-8) *Towards equitable decision making in smart cities*. RSE Conference.

NATIONAL INSTITUTE FOR HEALTH RESEARCH

Organ Quality Assessment

Colin Wilson

Newcastle upon Tyne Hospitals
NHS Foundation Trust



Challenge

Transplant is the best option for people with organ failure. Assessing organs for transplantation relies on the skills of the surgical team. Rate of use of organs varies widely across hospitals, between 30% and 70%.

Outcomes

Our tool will objectively evaluate organ quality and streamline decision making. Machine learning will be used to give organs a quality score prior to surgeon assessment. If used across the NHS it is estimated this tool could save hundreds of lives and millions of pounds through leading to 200 more kidneys, and 100 more livers being transplanted.

NATURAL ENVIRONMENT RESEARCH COUNCIL

PYRAMID

Liz Lewis
School of Engineering

Challenge

Floods present one of the highest risks to the UK. Flood models based on topology exist, but there are other significant factors such as blocked culverts or time of day that effect their likelihood and impact. For example, busy car parks or domestic rubbish bins out for collection can increase impact.

Outcomes

The RSE involved in this project has worked with the research team to support the development of multiple models describing some of the things that effect flooding. Data includes dynamic floating object information, rainfall and soil moisture and imagery of culverts and flood gates. These models will inform decision makers in near real time to make critical decisions during flood events.

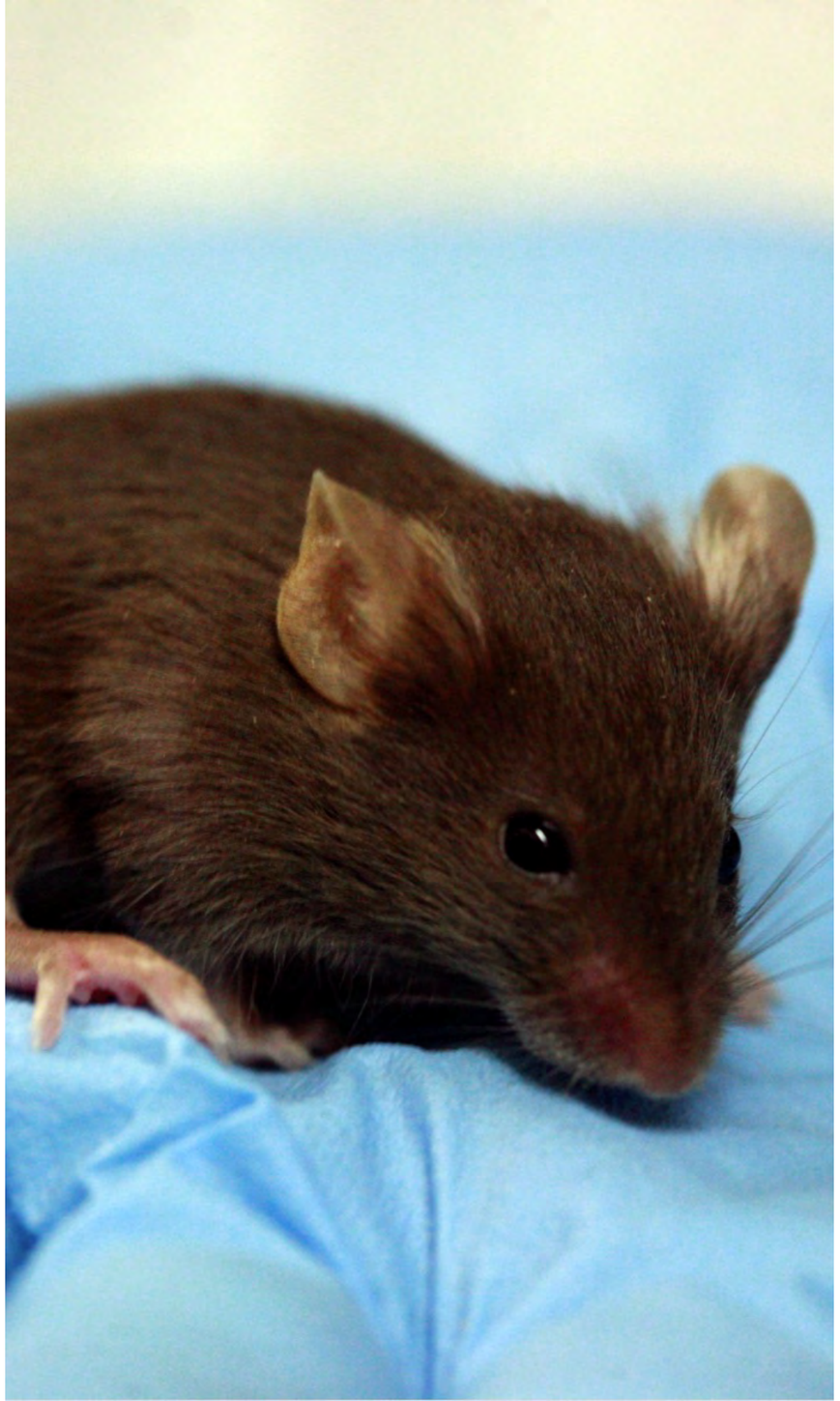
Wardle, R. (2022, July 5) *Analysis pipeline for hydrological simulation in the PYRAMID project*. DAFNI Conference.

NATIONAL CENTRE FOR THE REPLACEMENT REFINEMENT & REDUCTION OF ANIMALS IN RESEARCH

Mouse MApp

Matt Leach

School of Natural & Environmental Sciences



Challenge

Millions of mice are used worldwide in research each year (EU statistical report 2019). Traditional methods of assessing welfare based on changes in behavior or clinical signs (e.g. weight loss) are time consuming and can have other limitations as they may not be specific to pain or a sensitive indicator of health.

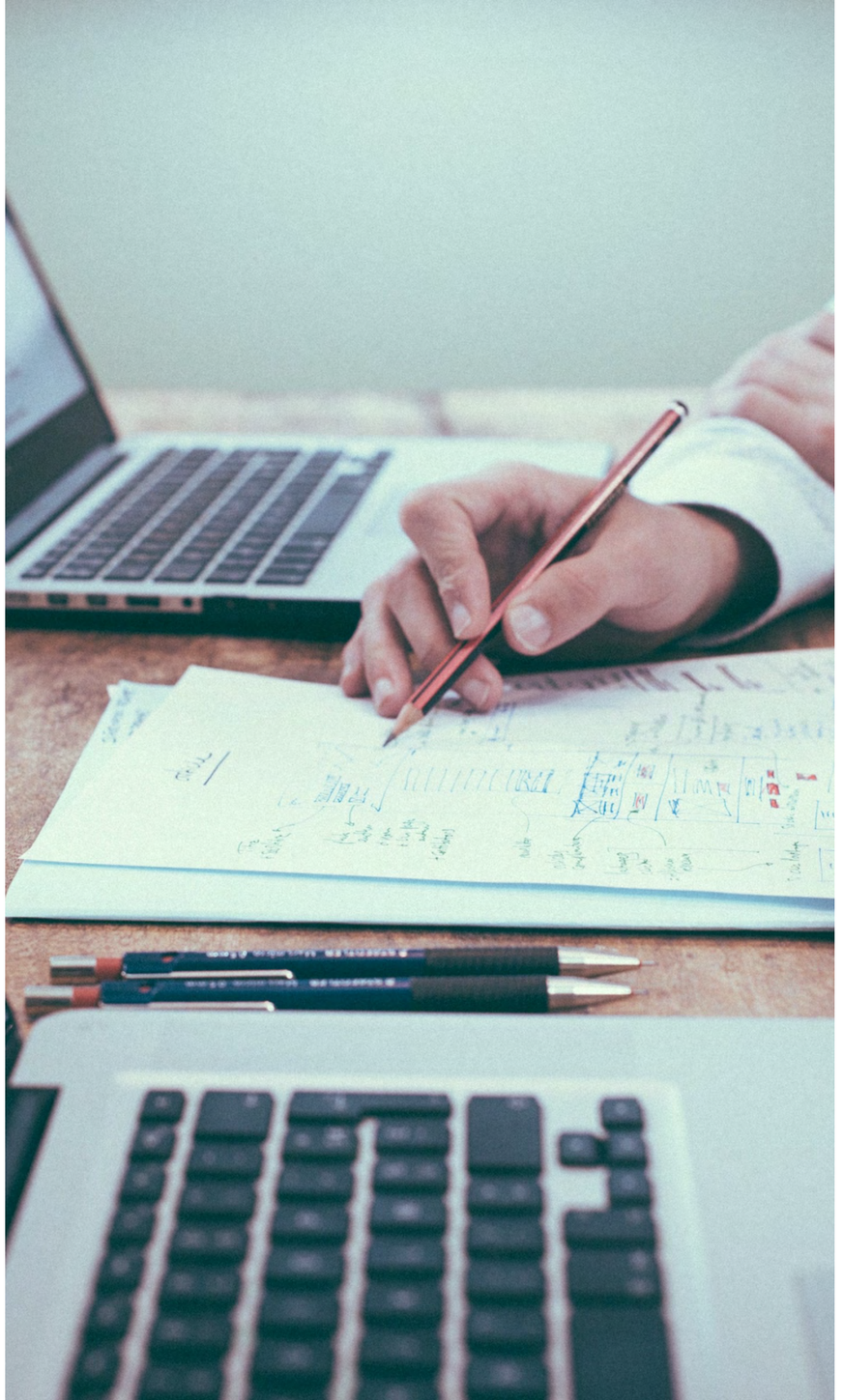
Outcomes

The RSEs involved with the project have created a classification algorithm that can estimate the health of a mouse from a photograph, placing it somewhere on a scale. Users interact with that model via a web app that is designed for use primarily on mobile devices.

NEWCASTLE UNIVERSITY

NUAcT Development Programme

Brian Walker
Deputy Vice Chancellor



Challenge

NUAcT Fellows often have not heard of the Research Software Engineering team; and may not be aware of how software engineering could support their research.

Outcomes

NUAcT Fellows were supported to upskill and develop funding applications involving RSE time through showcase events, Software Carpentries training sessions, and 1:1 support. The success of this programme has led to the rollout of a similar programme across the whole university.

REGIONAL STUDIES ASSOCIATION

Pandemic Effects

Rachel Franklin

Centre for Urban and Regional Development Studies

Challenge

It is speculated that the COVID-19 pandemic may have had a particular detrimental effect on female academics, as caring responsibilities increased.

Outcomes

The gender and location of primary and co-authors of academic journal articles was visualised to help researchers spot publishing patterns and understand hidden impacts of the pandemic.

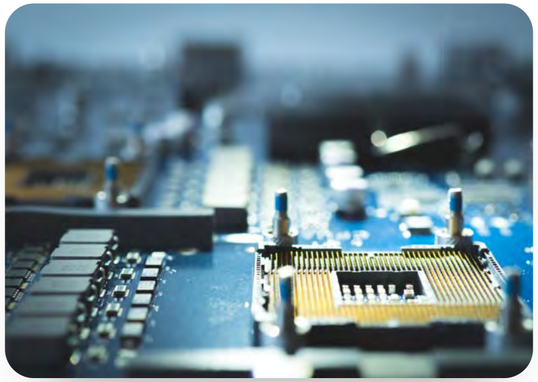
All Projects

Accelerated

Sean Wilkinson

School of Engineering

Plotting electricity substation data into GIS tools



ACHILLES

Ross Stirling

School of Engineering

Automated retrieval of sensor data from a simulated railway embankment at Nafferton Farm

Auto Generation of Deep Learning Networks

Steve McGough

School of Computing

Seeking more efficient ways of generating neural networks

Autosvar: Data Cleaning & Analysis Support

Emily Rainsford

School of Geography

Data analysis to help understand college graduate destinations in the North East

All Projects

Bee-ing Human

Jenny Richards

School of English Literature

Documenting interdisciplinary research inspired by Butler's 17th century book, The Feminine Monarchie of Bees

Chicken Shed Behaviour

Lucy Asher

School of Natural and Environmental Sciences

Exploring ways to use machine learning to identify when chickens exhibit dangerous swirling behaviours

Circular Chemical Economy

Aad van Moorsel

School of Computing

Tracking plastics as they move through the supply chain

Colouring Cities

Polly Hudson

Turing Institute

Developing software for platforms able to collect, collate, visualise and release open data on building stocks

All Projects

Cow Monitoring

Matt Leach

School of Computing

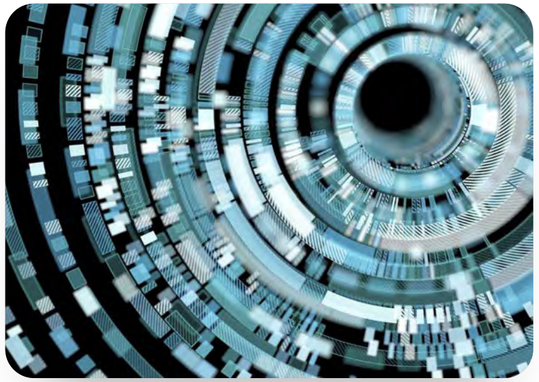
Using machine learning to assess the health of cows

CRITiCaL

Steve McGough

School of Computing

Run honeypot software on a range of cloud computing platforms in order to detect attempts to hack them



Data Safe Havens

Martin O’Reilly

Turing Institute

Exploring ways to store, handle and analyse identifiable information

DOLFIN

Jeremy Parr

Medical Sciences

Mobile app for the parents of neonates to submit clinical trial data

All Projects

EEG Data Processing

Yujiang Wang

School of Computing

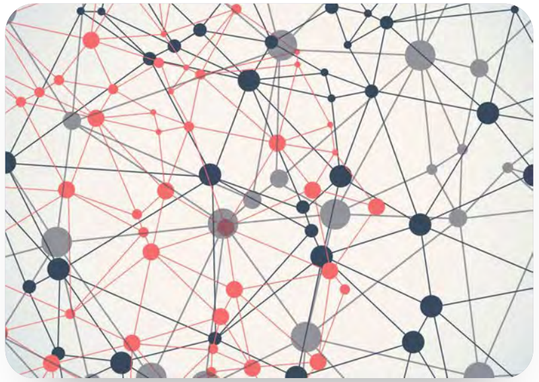
Implement a method for extracting spatio-temporal modes from high dimensional EEG data.

EpiChange

Peter Taylor

School of Computing

Quantifying longitudinal changes after epilepsy surgery



Evaluation Metrics for Synthetic Data

Steve McGough

School of Computing

Create a Python package that evaluates synthetic data against quality criteria

FinTrust

Aad van Moorsel

School of Computing

Explore the role of machine learning in banking, particularly in the context of automated lending decisions.

All Projects

Geosciences CDT Technician

Stuart Barr

School of Engineering

Providing specialist technical support for students studying on the Geosciences CDT

HDBR Database

Stephen Lisgo

Institute of Genetic Medicine

Application to allow the sharing of embryo tissue samples to different locations and research groups

Human Cell Atlas

Muzlifah Haniffa

Biosciences

Portal to allow the selection and display of specific features of human embryonic cells

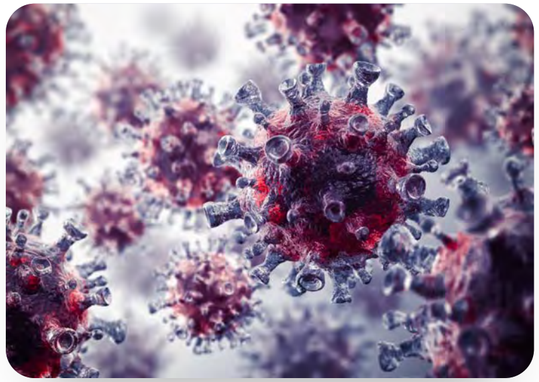
IDEA-FAST

Fai Ng

Translational and Clinical Research

Pipeline for fatigue data from wearable sensors and mobiles phones to an analysis platform

All Projects



Integrated Multi-Modal Tissue State Mapping

Boguslaw Obara

School of Computing

Developing data processing pipelines to integrate different modalities of cancer imaging for better prediction

Learning Machines

Alan Wilson

Turing Institute

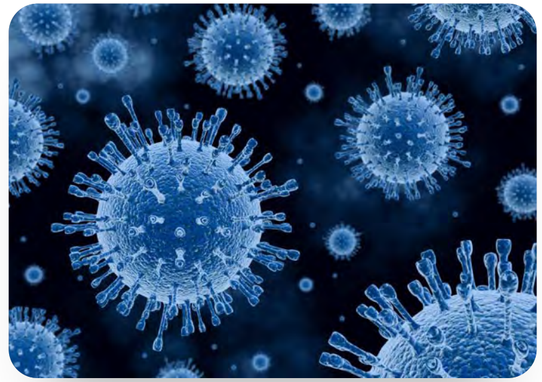
Developing infrastructure and methods for supporting continuous use of machine learning algorithms

Neural Networks for Spectroscopic Analysis

Tom Penfold

School of Natural and Environmental Sciences

Develop easy-to-use tools for the fast and automated analysis and prediction of X-ray spectroscopy



Newcastle Cancer Centre Pharmacology

Shelby Barnett

Translational and Clinical Research

Cell biology and clinical trials for healthcare research and cancer research training

All Projects

NHS Readability Tool

Gillian Rowlands

Medical Sciences

A tool that highlights complex sentences and words within a given text. It allows users to easily simplify text and lower the required reading age

NU Farms Data Organisation

Hannah Davis

School of Natural and Environmental Sciences

Creating a data pipeline for sensors

NU-LAB

James Cummings

School of English Literature

Developing a digital editions publishing cooperative that supports peer-reviewed digital scholarly works

Precision Diabetes

John Dennis

Turing Institute

Developing a decision support tool to optimise the selection of glucose-lowering therapy for people with type 2 diabetes

All Projects

PrivE

Callum Mole

Turing Institute

Evaluating the privacy of synthetic data with an adversarial toolbox

Refracted violence

Guy Austin

School of Modern Languages

Interactive digital journeys through archive of films documenting stories of civil conflict

Rheumatology Service Database

Arthur Pratt

Translational and Clinical Research

A database of clinical trial data from patients with rheumatoid arthritis

Rock Art in Oman

Stephanie Döpper

Goethe University

App where users can describe and log the condition of rock art examples in Oman

All Projects

Statistical & Computational Genetics

Heather Cordell

Population Health Sciences

Statistical methodology and software for determining relationships between genes and traits

SWEET

Linda Sharp

Population and Health Sciences

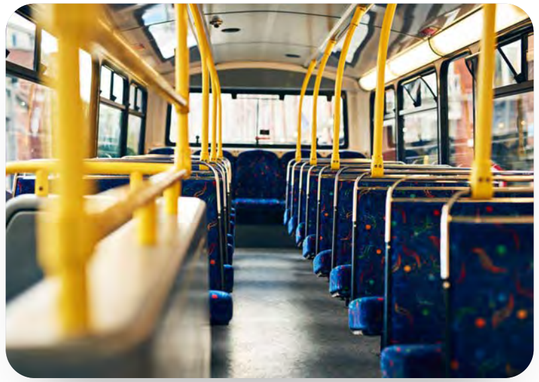
Support application for breast cancer patients to report their post-op medication symptoms

Talent development: Digital humanities

James Cummings

School of English Literature

Develop training in best practice within the digital humanities



TRACK

Paul Watson

School of Computing

Using machine learning to inform cleaning of public transport based on usage

Testimonials

Smart city technologies, particularly sensors, can potentially generate or reinforce existing socio-economic and spatial inequalities. By integrating sensor data streams, fine-scale demographic data, and movement data, this project aimed to create a decision support tool to help us think through who might be affected by potential sensor placements. Taking air quality sensors as a focus, the project developed algorithms to suggest possible sensor networks and evaluate their coverage for different population groups.

Research Software Engineers from Newcastle developed a website interface that supported decision makers in two ways: exploring population data; and viewing the algorithm-generated sensor networks on a map of the Local Authority. The intuitive and visually appealing interface provided a valuable first prototype for a decision support tool that can be used by a variety of stakeholders to evaluate the impacts of possible sensor networks in their area. The Research Software Engineers were central to project work, not only performing key software and analytic tasks but also serving as the project glue to bring together researchers from a variety of disciplines and backgrounds.

Spatial Inequality and the Smart City is a project funded by The Alan Turing Institute.

Rachel Franklin

Centre for Urban and Regional Development Studies

The team were very receptive to supporting us at short notice when we'd been let down by another provider. They worked with us in good time to create a fabulous app that is straightforward to use; and interacts with trial systems. The team worked with us to complete user testing; and made alterations. They supported us in the creation of a user guide. We have a functioning app for the start of our study and are really grateful for the team's expertise and support. We are now working with them on another initiative - and we'd definitely recommend them to others looking for software engineering solutions!

Jeremy Parr

Neurodevelopment and Disability

Challenges

Challenges

Looking ahead to 2023 there are several challenges facing the team, solving these will be a key part of how we grow in a sustainable way.

Legacy Code

Over time, RSE team members change, and project code becomes stale. There are proactive measures we can take to mitigate the impact of handling legacy code.

Forecasting Workloads

As our team and reputation grow, it becomes harder and more necessary to forecast our workload. This will help us manage commitments and recruit the right staff.

Service Expansion

When we were a small team, being agile was good as we could respond to change quickly. Now we need to think carefully through what our core offering is.

Legacy Code

Legacy code is source code that exists in older projects and may also be projects completed by former RSEs who have left the team. When projects are dormant or temporarily completed, these can be easily forgotten about. However, if they are then reactivated by a new burst of funding, they may need to be taken on by newer team members. This can present difficulties to the team member as they may be written in an unfamiliar language or have an unknown architecture. Unfamiliar systems can be difficult to understand. However, this can also be made worse by other problems.

Typical issues include:

- Older projects having out of date code libraries/dependencies leading to build errors. More modern versions of Node.js may not be able to run the code base, so switching to an older version is required
- Complex code structures may be insufficiently commented
- There may be a lack of sufficient documentation on essential system or deployment information such as project environment variables, keys, secrets, installation prerequisites, external systems APIs and so on
- There may have been insufficient handover information between team members leading to a loss of key information after staff have left

Some mitigation of legacy code issues can be to have the team follow the internal documentation process rigorously. Project markdown README files should be comprehensive and ideally all relevant (but not sensitive) files should be included in the project repository. Adequate project handover periods should be planned up to the end of staff contracts or leaving dates. Also, automated deployment should be used as much as possible, for example: Terraform for building cloud infrastructure and GitHub actions for deployment on pull requests. Problems around out of date code library dependencies can be difficult to resolve, so as much information as possible should be provided about project build instructions. GitHub can be configured to use Dependabot which is a tool that prompts repository owners to make security updates to dependencies through automated pull requests.

Forecasting Workloads

Forecasting how many staff we require, and with what skillsets, is challenging. The team has grown to accommodate our increasing workload, but under- or over-staffing presents a considerable risk. We may need to balance project types to reduce risk through more accurate forecasting.

Pre- and Post-award involvement
Involvement with Pre-Award funding applications means we can ensure costings are accurate. However, we cannot remove the uncertainty of whether a bid will be successful. Recruiting based upon the assumption of a positive outcome could leave us over-staffed. In contrast, Post-Award work brings more secure timescales and requirements. However, where researchers turn to the team to prevent underspend timescales can be challenging. We must manage expectations of what can be achieved within a fixed budget, particularly when RSE time runs up against hard deadlines from funders.

Project scale
Undertaking short projects mean RSEs become available frequently, ensuring we can be responsive to Post-Award, centrally funded and emergency projects. However, short projects can be demanding for RSEs if unanticipated challenges or periods of absence happen. Larger projects give security and lessen the administrative load of project management. Although the outcomes are proportionally bigger than shorter projects there is more potential to absorb issues and interruptions. There is a balance to be struck here between security and agility.

Team diversity
A key strength of the team is the variety of projects we can accommodate. It is rare we decline a project because we do not have the required skills. We have seen this year how forecasting what expertise the team needs is difficult as the number of projects requiring, for example, machine learning or web development, has peaks and troughs. Having a diverse portfolio is clearly a strength in attracting a larger amount of work, however it adds risks surrounding ensuring we have the right team composition.

We want to accept as many projects as we can for team security, to support research across the University, and to create a positive reputation. We do not decline work unless a project is not feasible (on technology or budget grounds) and this is rare. Indeed, we may take on projects even where the budget is challenging where we feel the teams' reputation will benefit, the researcher will bring in more work in the future, or we wish to support work with a public good. It would be detrimental not to have capacity for these edge cases, although they do place demands on the team. When to accept or decline work based on the impact on our workload forecasting detailed here will become pertinent as the team grows.

Service Expansion

A consequence of producing good work and having a reputation for delivering on projects is repeat collaborations as well as new interactions via word of mouth. This is obviously a good thing, but with more work comes more risk. If the processes and tools that we use don’t keep up with the pace of growth, then the quality of the output will suffer, and our reputation will be impacted. A challenge for the coming year is changing our processes to support that growth. The key will be doing that in a responsible way that maintains our standing, isn’t overly onerous to administer and keeps the team happy by letting everyone get on with their projects.

Carrying on as we have been with Excel spreadsheets and email chains to manage the state of each project is not sustainable, not now oversight of everything is required by more than one person. Some of the third-party services we use to track projects and timesheets are working well, and we’re investing development time on a management tool to that acts as a visualisation layer over all the data we collect from those services. It is important that whatever tools and services are being used that there is buy-in from the team and everyone agrees or at least understands the reason for their use. Without that buy-in, it can be an uphill struggle to get people to use them and capture the data we require to keep the team running.

In the early part of the year we underwent a restructure to bring in more management resource, a large driver of which was looking ahead to improving our processes. Whilst that restructure was not an easy time, it was done as openly as possible so that everyone felt involved and informed throughout. This spirit of openness is something that has been part of the team management style from the very beginning, and this will continue as we make tweaks and changes to how we do things in the year ahead. There will be regular updates in our team meetings, but most crucially all our decisions and updates will be made on GitHub so that the entire team can view what's going on at any time and contribute to any proposed changes. Everyone in the team appreciates that we have outgrown some of our current ways of doing things, the hard part is going to be changing the right things in the right way and not disrupting what works well.

Opportunities

Opportunities

Whilst we are experiencing some challenges, there are also many opportunities to expand our service offering as we grow.

Internal Processes

As the team expands, there is a need for better management of the project lifecycle. We are changing how we work to strengthen and add to our existing processes.

Scaling Training

Training researchers in software engineering best practice benefits the University, research and the RSE team. A larger team and Operations Assistant enable us to scale our offer.

Research Computing

There are opportunities to contribute to digital research beyond purely software. We should be looking at hardware, dev-ops and building a community of practice around each area.

Processes

With growing levels of project experience, we have discovered the need to focus more on the management side of software development. The following processes help counter some of the problems with legacy code and create a more organised approach to project management generally. Key areas include:

- The correct initial scoping of projects - time estimation on software development can be tricky, however allowing enough time for the completion of requirements is essential. We are starting to involve a wider number of team members with project scoping so that specialist knowledge is brought in at an early stage.

- Capturing key information such as stakeholders, deadlines, project importance, etc. We are looking at various options for recording and sharing information including project starter forms, checklists that can be used as discussion prompts and documents that set out our terms and conditions.

- The entry of project information into HubSpot – this includes contacts, project proposals, emails and notes. This information is then centralised rather than being spread across emails, SharePoint, personal drives and spreadsheets.

Ideally, there should be an ongoing and rigorous documentation process throughout the project, so that information is timely and relevant. This should not be left until the project end. During the project, RSEs should have regular conversations with PIs and other stakeholders. Managers should take a pro-active approach to following up with PIs during staff transition periods.

In general terms, testing is essential on all but the shortest projects. Automated deployment is also always preferable, but if a project is so short that this is not practical, then thorough documentation on the deployment process should be supplied.

Finally, the team is moving towards a greater reliance on GitHub and its tools. The team are able to implement GitHub best practices, for example: protecting their repo main branch if appropriate, using feature branches and managing issue progress through the project's facility. We have recently created a template repository for all new projects that contains user story, task issue and pull request markdown templates, so that we have more consistency of process.

Training

In 2022 we expanded our training to deliver 7 Software Carpentries workshops each lasting 2 or 3 days. Most covered the core Software Carpentries syllabus of UNIX command line, Git and Python, with 2 bespoke courses using other Carpentries material. This increase was made possible by our dedicated Training Lead, and the expanding number of fully trained Carpentries instructors within the team. Looking forward at the next year, the ambition is to deliver 10 workshops, with logistical support from our new Ops Administrator. With capacity for 20 participants in each workshop, we could reach up to 200 postgraduate students and researchers across the university.

Providing in-house training in core software engineering skills to this audience has clear benefits. Reproducible, efficient and accurate science is underpinned by these skills, and they are transferable across domains. Whilst some projects require the depth of knowledge and sustained effort of an RSE, others can be supported by upskilling the research team in software engineering. Delivering training benefits the team by extending our impact, developing a baseline of programming skills across the university, and raising our profile. Receiving training from the team benefits the university through providing CPD without the cost of an external provider, increasing standards, and training the next generation of researchers in best practice.

There are several opportunities within the teaching that are yet to be leveraged. Our Training Lead is certified to train others as Carpentries Instructors. This will allow us to train more people across the University and more quickly train new RSE team members. More instructors leads to a second opportunity, increasing the number of workshops and range of content. As the number of ‘graduates’ of the core syllabus increases, there is likely to be a number who wish to progress to an intermediate and expert level. Offering more complex courses would allow us to also reach researchers who already write code; but could benefit from learning software engineering principles they might not have been exposed to. The final opportunity is to further consider who we train. We currently advertise through workshops.ncl and workshops are consistently over-subscribed. However, the majority of our bookings come from postgraduate students. Whilst this is a valuable population to train, we may also secure additional impact by targeting research staff. This may be addressed through our tour of the university in 2023 as more researchers become aware of our training. Of course, these routes to expansion require staffing. Whilst our target of 10 workshops in 2023 seems achievable, we may need to explore additional resourcing for training should we decide to increase this number or expand our syllabus in the future.

Research Computing

Support for researchers needing access to digital skills, tools and platforms is patchy. Currently researchers access different services and units for different needs, NUIT supports HPC provision and runs the research data storage service and the RSE team provides custom software engineering and data science. No one is formally supporting access to Cloud Computing or Secure Computing and access to grant writing assistance, 1-to-1 support and training across all digital research is severely limited. The consequence of these barriers is less high-quality research, fewer impact case studies, and lower research income.

A solution to this that has seen success at other institutions is the create of service and/or research units for the tools, practices and systems that enable digital research. The unit could have both technical services and academic research missions, delivering digital tools and research platforms as a service and conducting research into the application of computational and data science methods. A comprehensive training and community support programme would raise the base-level skills of researchers and the training material could ultimately be worked into taught modules for degree programmes.

Building on the success of the RSE work we do, there are four broad areas that the unit could lead in

Expertise
A pool of technical staff covering activities such as Software Engineering, Data Science, Data Management and Research Infrastructure.

Training
Maintenance and support for research platforms big and small, from IoT to Exascale-HPC and everything in between.

Platforms
A suite of training courses aimed at all career stages and experience levels covering all aspects of research computing.

Support
Community support via extensive tutorials, guides, drop-in sessions, technical talks and community events.

Newcastle University

Research Software Engineering

Catalyst Building
Room 2.01
Newcastle University
1 Science Square
Newcastle Helix
Newcastle upon Tyne
NE4 5TG