

# Research Software Engineering

2019

# Contents

3	Overview	55	Testimonials
13	Finances	59	Challenges
20	Team Members	65	Opportunities
30	2019 Projects	71	2020 Projects
49	Outreach	82	Appendices

# OVERVIEW

# Introduction

---

The concepts of software engineering best practice have been central to the Digital Institute's research activities since its inception in 2010. In that time several members of staff have been creating software for a range of research projects. Since 2016 there have been efforts to have the role of Research Software Engineer (RSE) recognised at UK research institutions [A]. The Digital Institute has been driving this recognition here at Newcastle, culminating in the decision to create an RSE team in late 2018.

The first year of the team's activities has seen rapid growth both in staff and project partnerships. In the first instance, the initial round of recruits had not arrived in post before we went out to advert again for an extra three RSEs. In the case of projects, these have grown organically through existing relationships and word of mouth, with little need for marketing the team's services.

2019 has shown that the University's decision to create an RSE team has already paid off, with demand for its services far outstripping supply. Balancing that demand with staff availability will be paramount to a successful year in 2020. The team faces a number of challenges connected to its rate of growth, but also exciting opportunities to contribute in other areas of the university. The team now has a project income of £1.4m from 28 different projects with a total research value of £30.4m.

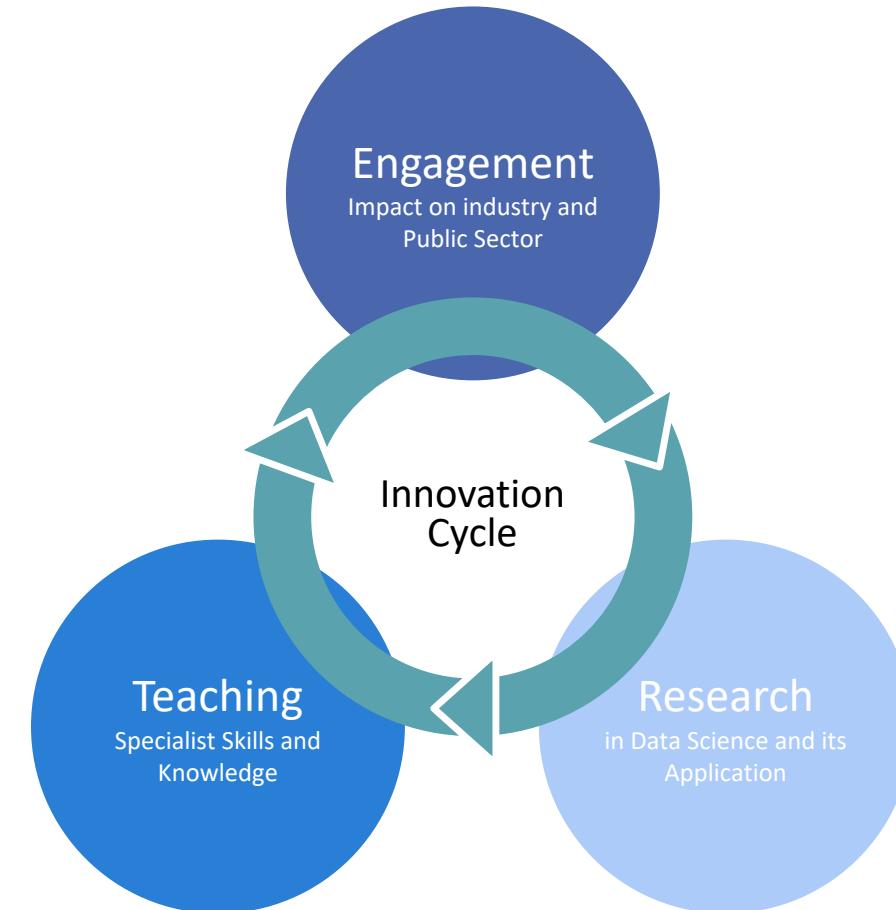
Mark Turner  
Head of Research Software Engineering  
Digital Institute  
Newcastle University

[mark.turner@ncl.ac.uk](mailto:mark.turner@ncl.ac.uk)

# Newcastle Data

---

The RSE team sits within the research arm of Newcastle Data. Newcastle Data fulfils three core functions; teaching the next generation of researchers and data scientists, engaging with industry and the public sector and traditional research activities. The RSE team plays its part across all three but is primarily focused on supporting research.



# The Team Model

---

The team operates on an internal consultancy model, charging its staff onto research grants. With the exception of Mark Turner and another half of a post, 100% of direct salary costs for the whole team must be recovered. This is done by costing individuals onto grants.

**POOLING RESOURCE**  
Sharing resource across projects reduces bus-factor and ensures continuity of support for the research project



**FLEXIBLE ALLOCATIONS**  
RSEs can be assigned to projects at key times on the project, and not always continuously



**SKILL SHARING**  
Projects benefit from fractional FTE splits of RSEs with different skills and experiences



## Review of

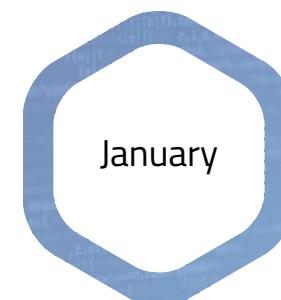


**2019**

The first year of the Research Software Engineering team has been one of rapid growth and increasing demand for software engineering consultancy across the entire University.

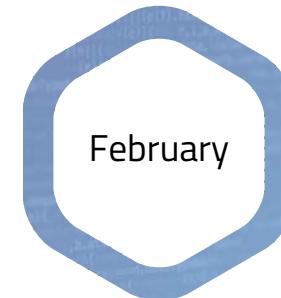
## RECRUITMENT & PROJECTS

After interviews, offers are made to four candidates.  
Mike Simpson transfers into the team.  
Pancreas Transplant Imaging project starts.



## SHORTLISTING

The recruitment campaign started in late 2018 closes.  
18 candidates apply and 11 are shortlisted for interview.  
Geosciences CDT Technician added to project list.



## TURING SPF PROJECTS & RECRUITMENT

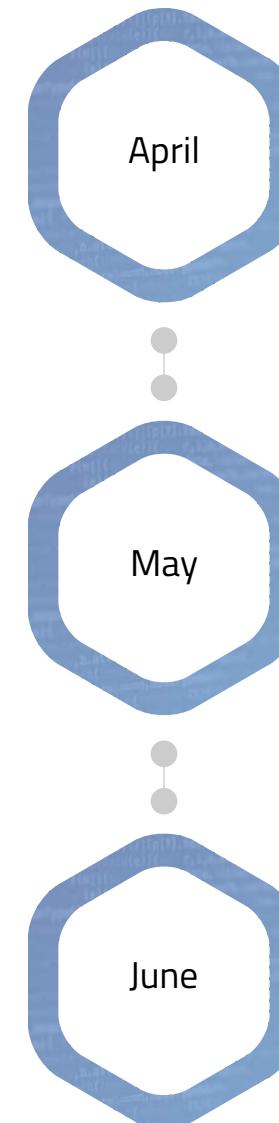
The Alan Turing Institute launches SPF themes [B].  
Digital Health event in Manchester and TPS in Cambridge.  
Recruitment for three more posts starts.

## RECRUITMENT & COMMUNITY

12 Candidates apply and 6 are shortlisted for interview.

Team attends Thinking Digital 2019.

Mark Turner becomes inaugural trustee of Society of RSE.



## TEAM LAUNCH

Team grows to 6 people.

Kate Court, Fiona Galston, David Herbert & Jannetta Steyn join the team.

Visit to Slayley Hall for a two-day intensive on-boarding event..

## TEAM EXPANSION

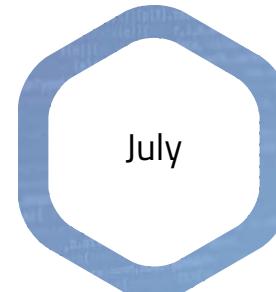
Team grows to 9 people.

Andy Hardy, Frances Hutchings & Rebecca Osselton join the team.

ePraSE 2019 v1.0 Released.

## NEW PROJECTS

RIF HaSS Pilot Projects call opens.  
Agreement to work on FinTrust project.  
Agreement to work on FRONTCAT project.



## CONFERENCE SEASON

RSEs working on Turing projects visit British Library.  
Mark & Fiona attend Digital Humanities 2019 in Utrecht.  
Agreement to work on MISTRAL project.

## RSE CONFERENCE

The whole team attends RSE Conf 2019 in Birmingham.  
Hate Speech project starts.  
Agreement to work on Human Cell Atlas project.

## MORE NEW PROJECTS

Agreement to work on QMOD Atlas.

Agreement to work on Rome Transformed.

Recruitment starts for two posts.

October

November

December

## PILOT PROJECTS

The first round of HaSS Pilot Projects start.

MISTRAL project starts.

Lucy McLaughlin joins as a temporary assistant.

## RECRUITMENT

20 candidates apply and 4 are shortlisted for interview.

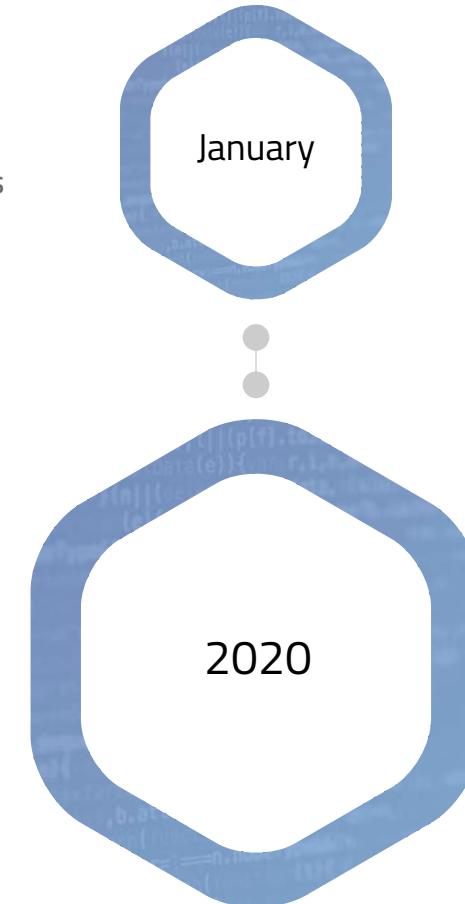
Digital Institute transitions into Newcastle Data.

RSE Team moves into the Catalyst Building with NICD.

## EXPANSION

Team grows to 11  
Human Cell Atlas project starts  
FinTrust project starts

## Outlook for



Next year will need to be a year of consolidation after such rapid growth. Identifying projects and funding sources that allow the team to retain staff and secure a strategic role within Newcastle Data will be the focus.

# FINANCES

# Overview

As the RSE team is based within Newcastle Data, which is within SAgE, the bulk of our projects have come from that faculty [C]. However, there is growing interest in our services right across the university which should lead to significant growth in all faculties in the coming years.

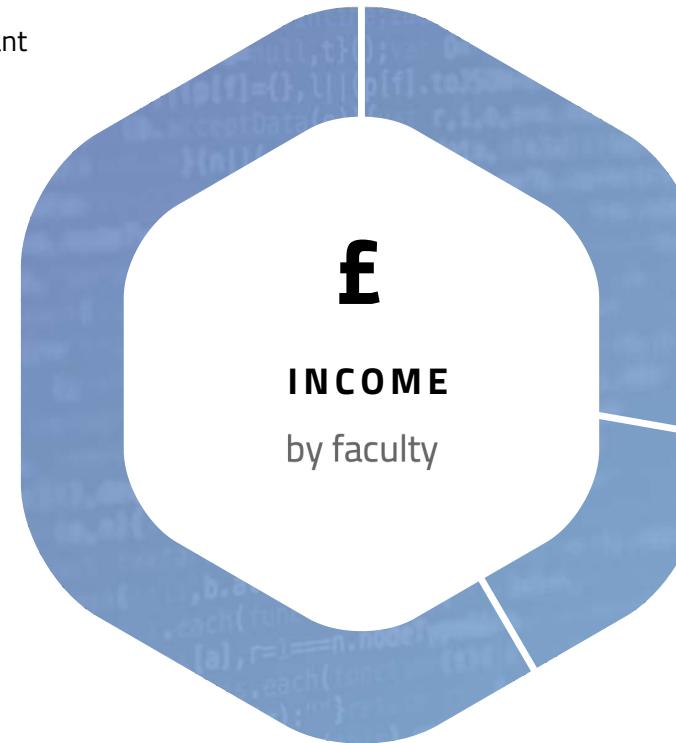
## SCIENCE, AGRICULTURE & ENGINEERING

14 Projects    £681,061

**Projects**    **Income**

60% of Income    £24,770,094

**of Income**    **Research Value**



## HUMANITIES & SOCIAL SCIENCES

7 Projects    £327,029

**Projects**    **Income**

29% of Income    £2,663,584

**of Income**    **Research Value**

## MEDICAL SCIENCES

5 Projects    £127,235

**Projects**    **Income**

11% of Income    £11,183,442

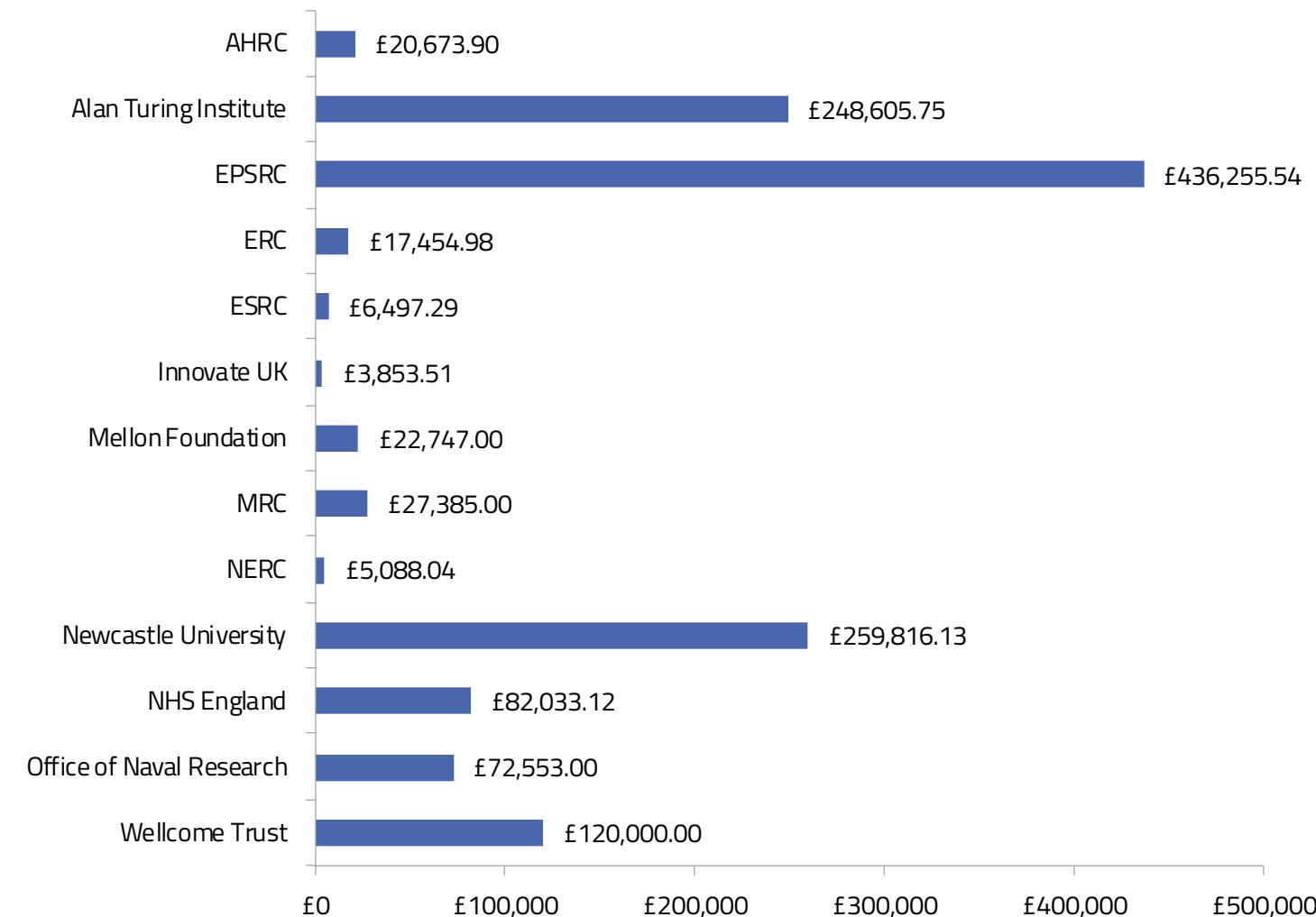
**of Income**    **Research Value**

**FINANCES**

# Funding Sources

---

The distribution of funding sources mirrors the breakdown by faculty. 60% of our work is in the SAgE faculty so it makes sense that some of the largest funders are EPSRC and the Alan Turing Institute.

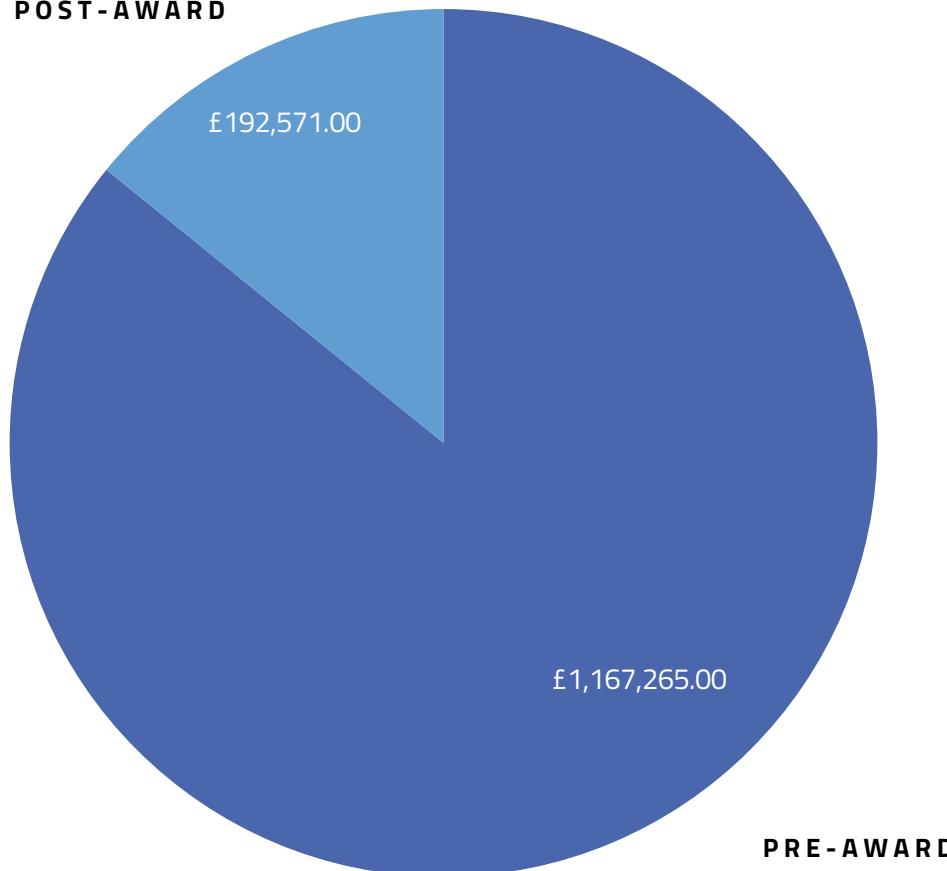


**FINANCES**

# Award Stage

---

Despite post-award work accounting for 43% for our projects, it accounts for only 14% of our financial income. This highlights the importance of being costed onto grant applications to give some long-term security to team members.

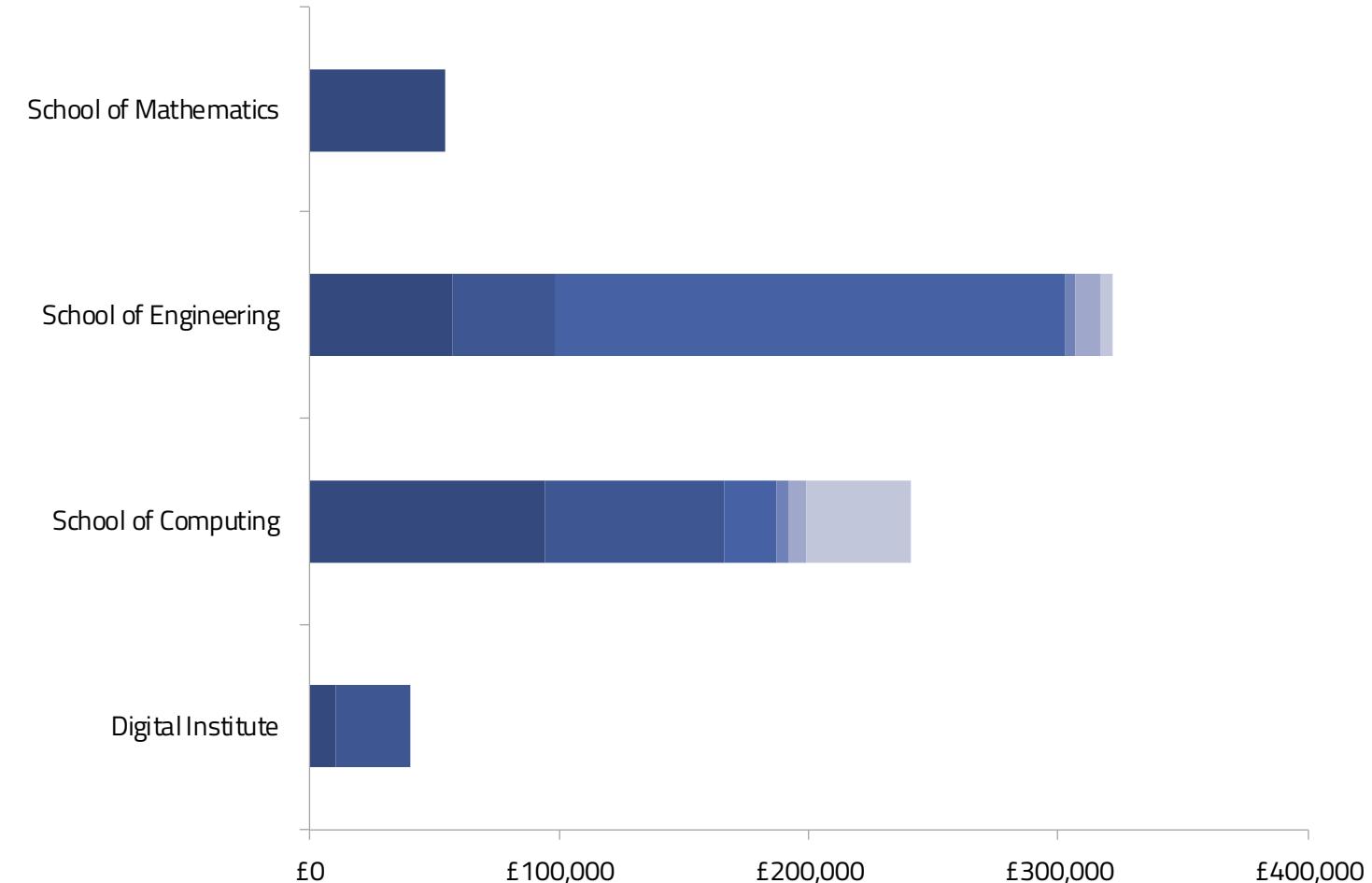
**POST-AWARD**

## FINANCES

# SAgE Income

---

As previously mentioned, being within the SAgE faculty means that the bulk of our connections and day-to-day access to researchers come from this faculty. It's no surprise to see larger income as a result.

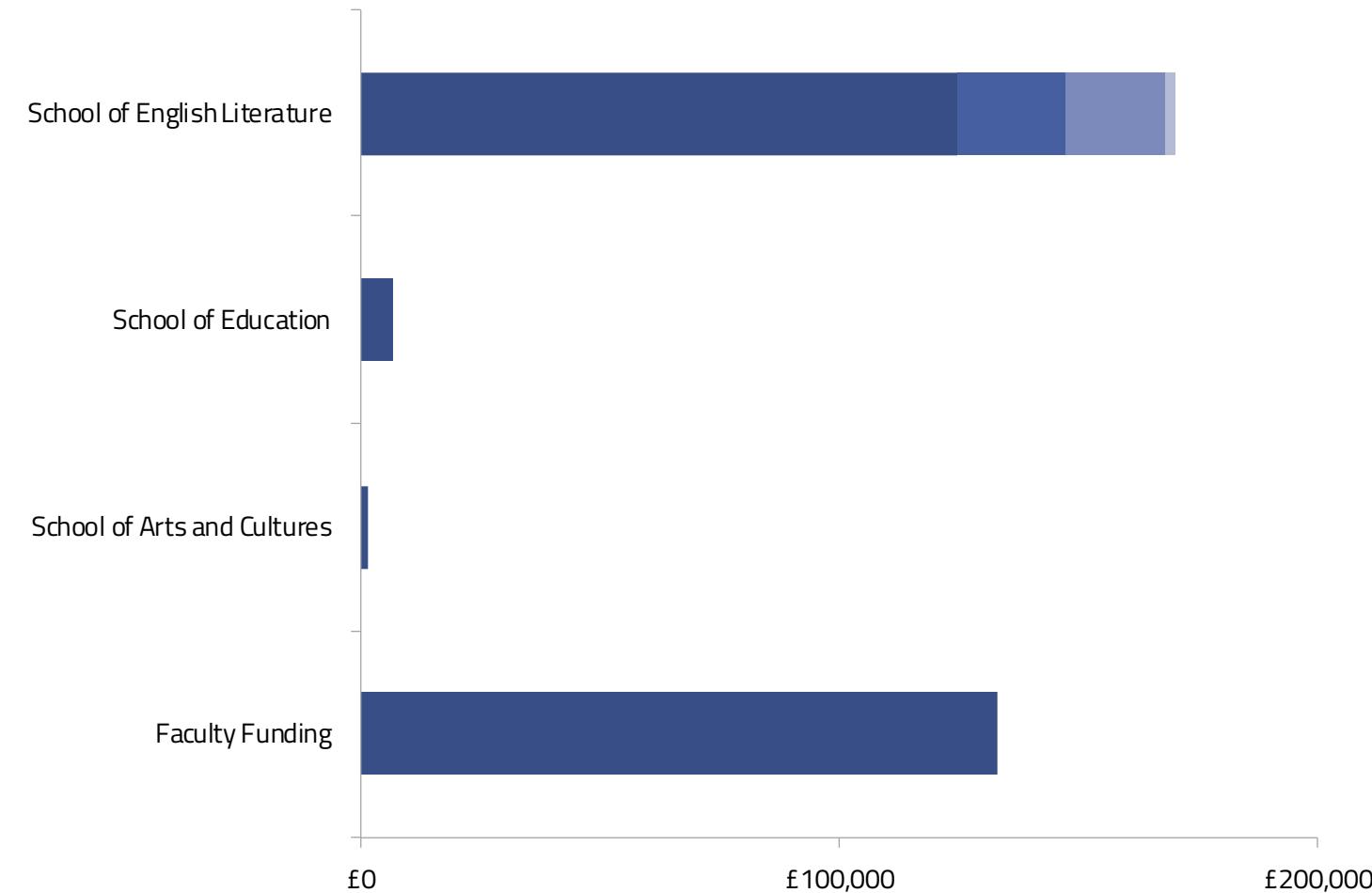


## FINANCES

# HaSS Income

---

HaSS is one of our most important strategic partnerships, with two flagship projects: ATNU with Jenny Richards and the RIF funding for dedicated support of HaSS researchers.

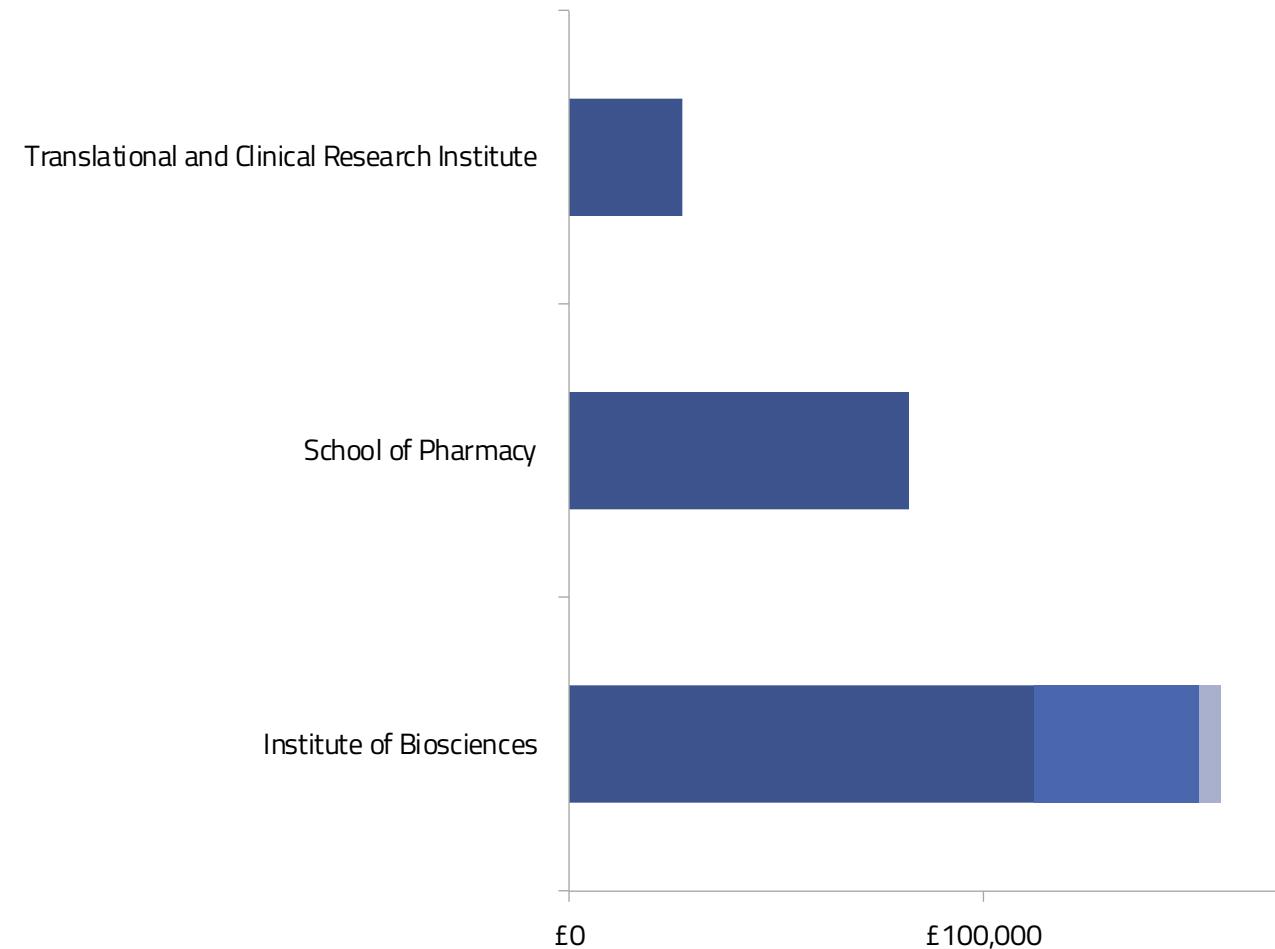


## FINANCES

# FMS Income

---

FMS has been undergoing its own restructure in the last year, so projects there are now split across different new institutes. FMS is a key area for outreach in 2020 to increase our presence there.



# TEAM

# Kate Court



Fine Art

**Background**

Web Applications & Middleware

**Key Skills**

01/04/2019

**Joined**

Kate joined the team from Northumbria University where she was working as the Project Development Manager on the Creative Fuse project. She had recently completed the MSc Computer Science conversion course at Newcastle University. Before that Kate completed a PhD in Fine Art at the University of Edinburgh.

Fiona  
**Galston**



Computer & Network Technologies

**Background**

Full Stack Web Development

**Key Skills**

01/04/2019

**Joined**

Fiona joined the team from Accenture where she had been working in a team supporting clients in the nuclear energy and healthcare sectors. She graduated from Northumbria University, switching to computer science after starting a degree in biology at the University of Manchester. Fiona is a full stack web developer with knowledge of a range of JavaScript frameworks.

# Andy Hardy



Computer Science

**Background**

Web Applications & Cloud Computing

**Key Skills**

03/06/2019

**Joined**

Andy joined the team from DEFRA, having previously worked as a contractor at several well-known North East technology firms. Andy brings 20 years' worth of experience of software development in the private sector with a wide range of programming languages and frameworks.

David  
**Herbert**



Natural Sciences

**Background**

Geospatial Systems & Web Services

**Key Skills**

01/04/2019

**Joined**

David joined the team from the British Antarctic Survey where he had been a developer for over 10 years. He brings a wealth of knowledge of geospatial systems, web services and web development more broadly. David is passionate about open source and has been an active participant in several communities.

# Frances Hutchings



Computational Neuroscience

**Background**

Machine Learning & Data Modelling

**Key Skills**

03/06/2019

**Joined**

Frances joined the team from an RA position in the ICOS research group in Computing. Frances specialises in data analysis and modelling and is applying these skills to machine learning problems. Frances brings knowledge of the challenges of working with large datasets and the application of HPC techniques, and programming languages such as MATLAB, Python and R.

# Rebecca Osselton



Computer Science

**Background**

Web Applications & Digital Healthcare

**Key Skills**

03/06/2019

**Joined**

Rebecca joined the team with many years' worth of experience working in both the public and private sector. She brings invaluable experience of working within the NHS from a previous role. Rebecca is a full stack web developer competent in a range of PHP and JavaScript frameworks. She has an interest in supporting young people and women to engage in the technology sector.

# Mike Simpson



Computer Games Engineering

**Background**

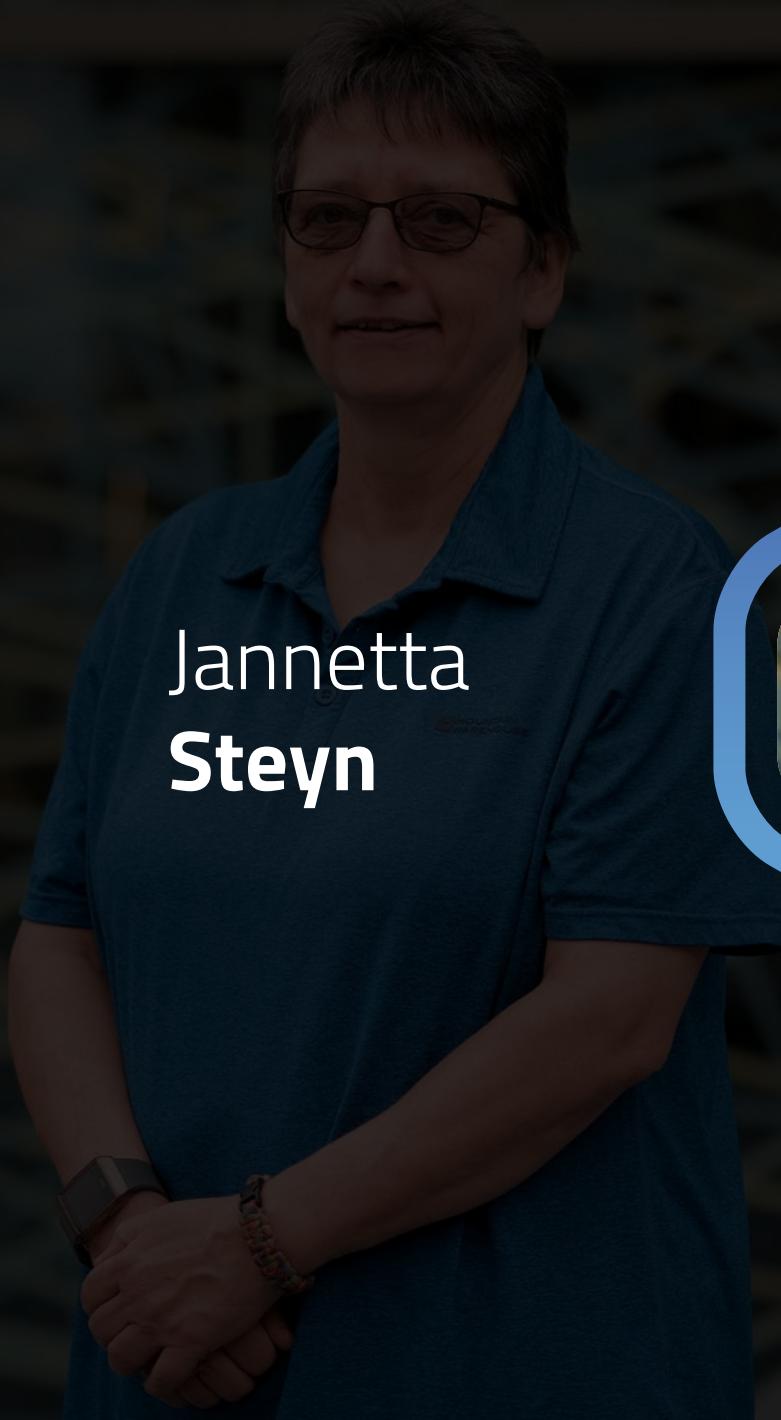
Visualisation

**Key Skills**

01/01/2016

**Joined**

Mike completed a BSc and MSc at Newcastle, then worked as a Game Programmer before returning to university to study for a PhD. He is a generalist programmer with experience in gameplay, graphics, physics, web and interface programming in a range of languages. His interests include photography, music, writing fiction and creating digital artwork.



Jannetta  
**Steyn**



Computational Neuroscience

**Background**

Bioinformatics, Provenance & Middleware

**Key Skills**

01/04/2019

**Joined**

Jannetta is an experienced researcher and software engineer who has held a number of RA and senior RA positions at the University. She has a background in data analysis, provenance and middleware programming. Previously she has worked in the Bioinformatics Support group in the Medical School. Jannetta does a large amount of outreach work, primarily in STEM, running a range of coding clubs and electronics clubs.

# Mark Turner



Computer Science

**Background**

Web Applications & UI/UX

**Key Skills**

01/04/2019

**Joined**

Mark joined the University in 2012 and has worked as Software Engineer in the Digital Institute for the last 8 years. He worked on a number of large research projects during that time, specialising in web applications and user interface design. Mark was a founding trustee of the Society of Research Software Engineering and is an active member of national and international RSE communities.

# 2019 PROJECTS



# Animating Text Newcastle University

---

Animating Text Newcastle University (ATNU) is a digital collaboration between scholarly editors based in humanities disciplines and the Digital Institute. It sets out to create new ways in which readers and users can interact with texts, and to explore and test opportunities for immersive reading and writing. What's unique about ATNU is that our ideas for the immersive texts of the future are based on the texts and books of the past that we are editing (1500-1900), which were already imagined as variable, dynamic, vital, interactive, akin to a 3D experience.

Jenny  
Richards

**Principal Investigator**

Humanities & Social  
Sciences  
**Faculty**



# Automating Data Visualization

---

The project proposes to investigate novel psychophysically informed algorithms to automate the visualisation process. This will address a growing gap in data science, the link from data and analytics results to human cognition and decision making. It is evident that, despite advances in AI, without a better human comprehension of data derived knowledge and its related uncertainty we stand little hope of improving the collective sum of human wisdom. Currently, the scale of big data is challenging the capacity of visualisation designers to bridge this gap. The aim is to exploit the power of cloud computing to automate the process of laying-out and annotating a 3D/4D visualisation to best fit a desired visual outcome.

Nick  
Holliman

**Principal Investigator**

Science, Agriculture  
& Engineering  
**Faculty**



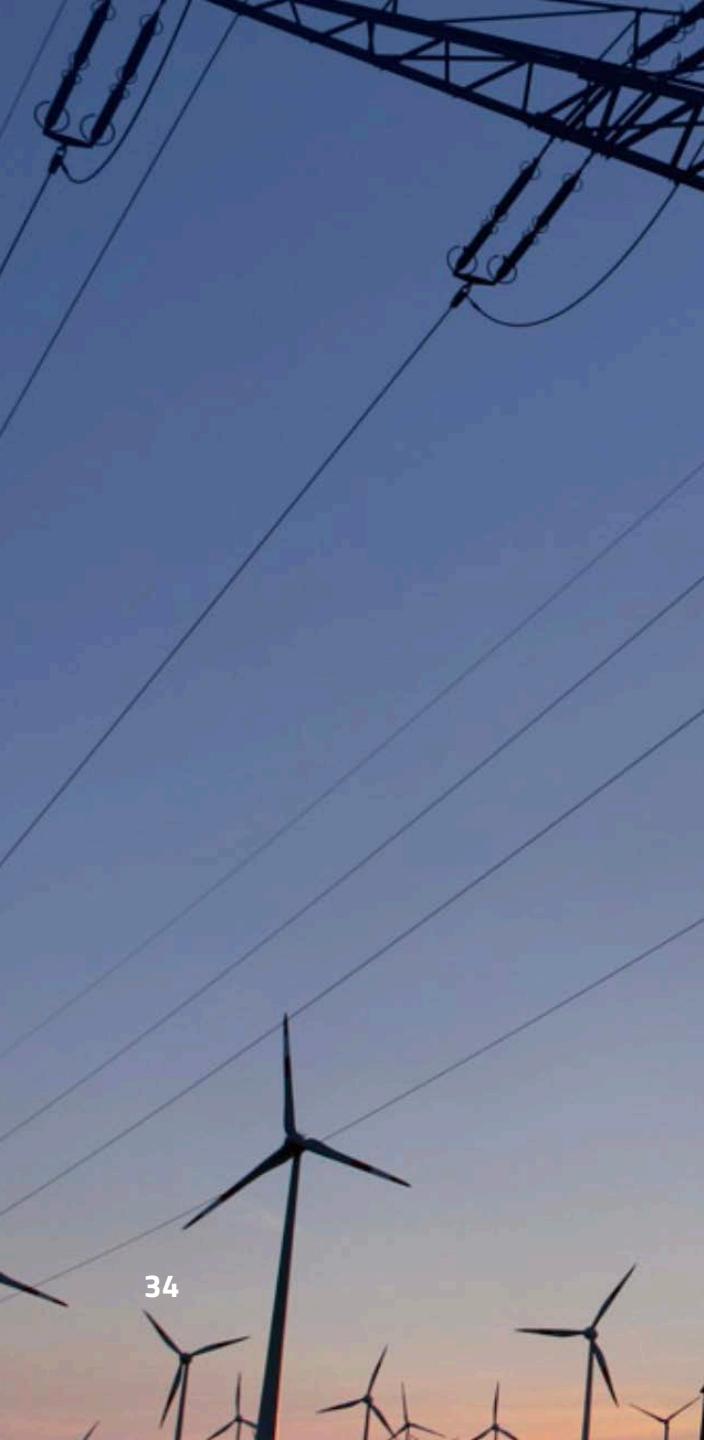
# Controlling Abnormal Network Dynamics using Optogenetics

CANDO (Controlling Abnormal Network Dynamics using Optogenetics) is a world-class, multi-site, cross-disciplinary project to develop a cortical implant for optogenetic neural control. The goal is to create a first-in-human trial in patients with focal epilepsy. In the brain, nerve cells generate rhythmic activity or 'brain waves'. In many neurological diseases these rhythms are disrupted, producing abnormal patterns of activity. In epilepsy, abnormal activity can often be localised to a small 'focus', which then spreads causing a seizure. CANDO proposes a treatment using a small implant to modulate abnormal activity and so prevent seizure development. The implant provides precisely timed stimulation by continuously monitoring brain waves via implanted electrodes and modifying them via implanted light sources.

Andrew  
Jackson

**Principal Investigator**

Medical  
Sciences  
Faculty



# Centre for Energy Systems Integration

---

34

The Centre for Energy Systems Integration (CESI) aims to reduce the risks associated with securing an integrated energy system for the UK. CESI is a unique partnership of five research intensive universities and strategic industrial partner Siemens. The energy trilemma presents many complex interconnected challenges. Current integrated energy systems modelling and simulation techniques suffer from some shortcomings, so models are unable to provide accurate or detailed enough integrated representations. The aspects of real energy systems aren't taken into account sufficiently and they also struggle to generate robust long-term plans in the face of uncertainties. CESI aims to address these challenges by adopting a multi-vector and multi-vector approach within a highly collaborative environment.

Phil  
Taylor

**Principal Investigator**

Science, Agriculture  
& Engineering  
**Faculty**



# Communications & Trust in Emergencies

---

Communication in Emergency Management through Data Integration and Trust (CEM-DIT) is about getting information about the state of a disaster to people who have to make decisions on disaster response, in a way that allows them to understand and trust that information. The aim of the project is to provide an automated system to support decision makers during emergency responses by providing them with the information they need in a timely and usable fashion. Is it possible to integrate matching technology with provenance-aware security in order to allow fast, effective querying of mismatched data sources, and to annotate the response such that the querier is able to understand its quality and meaning?

Paolo  
Missier

**Principal Investigator**

Science, Agriculture  
& Engineering  
**Faculty**



# Comprehensive Assessment of Reading in Aphasia

---

The Comprehensive Assessment of Reading in Aphasia (CARA) has been designed to be used by Speech and Language Therapists (SLTs) and other qualified professionals to assess reading in people with aphasia. CARA was designed with and for people with aphasia, and SLTs. It provides a comprehensive profile of someone's reading, with linguistic sub-tests examining single word, sentence and paragraph level comprehension. The RSE Team worked with the developers of the test to produce a web-based version that collates and summarises responses to create a report immediately on completion. This report can then be used by the SLT when planning further assessment and treatment.

Julie  
Morris

**Principal Investigator**

Humanities &  
Social Sciences  
**Faculty**



## e4Futures: Vehicle Grid Integration

---

Electric vehicles (EVs) can break our dependence on fossil fuels in the transport and energy sectors. However, mass adoption of EVs introduces significant and disruptive electricity demand to meet the charging needs of these vehicles. Vehicle grid integration strategies, underpinned by data science, ensure that electric vehicle charging infrastructure is synergistic with the electricity grid, reliable, cost effective and sustainable. This work can be applied in the automotive and energy industries to help empower companies to make informed investment decisions and develop products which are future-proof and user-centric. The work can also provide evidence to support informed energy and transport policy decisions.

Myriam  
Naimeah

**Principal Investigator**

Science, Agriculture  
& Engineering  
**Faculty**



# ePrescribing Risk and Safety Evaluation Tool

---

The ePrescribing Risk and Safety Evaluation (ePRaSE) project seeks to create 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. The ePRaSE tool seeks to test these rules by creating patient scenarios that test the ability of ePrescribing systems to flag issues as well as let through legitimate and safe prescriptions. It captures and compares data from all trusts to provide an overview of the state of ePrescribing in NHS England.

Stephanie  
Klein

**Project Manager**

Medical  
Sciences  
Faculty



# FRONTCAT: UI for the CityCAT Flood Model

---

City Catchment Analysis Tool (CityCAT) is a software tool for modelling, analysis and visualisation of surface water flooding. CityCAT enables rapid assessment of combined pluvial and fluvial flood risk. It allows assessment of the effects of different flood alleviation measures using a combination of factors such as the use of standard, readily available data sets and sophisticated and efficient algorithms for grid generation. The FRONTCAT project seeks to create a user interface over a cloud native version of the original CityCAT application.

Chris  
Kilsby

**Principal Investigator**

Science, Agriculture  
& Engineering  
**Faculty**



# Geospatial CDT Technician

---

The CDT is an integrated collaboration between Newcastle University and the University of Nottingham, two world leading centres of Geospatial Systems research. Drawing on experts from across Newcastle University, the CDT is a unique opportunity for academia to work with the geospatial industry to address major global societal issues, such as climate change impacts, urban sustainability, spatially resourcing public health, and removing spatial barriers to social inclusion and healthy ageing. The RSE team support this effort through a dedicated resource for software development advice and guidance to students, ensuring code produced from the CDT is of a high standard for reuse in the geospatial community.

Stuart  
Barr

**Principal Investigator**

Science, Agriculture  
& Engineering  
**Faculty**



# Hate Speech: Measures & Counter-Measures

---

Hateful content online is a growing problem in the UK. It can pollute civic discourse, inflict harm on targeted victims, create and exacerbate social divisions, and erode trust in the host platforms.

This project is developing and applying advanced computational methods to systematically measure, analyse and counter hate speech across different online domains, including social media and news platforms. The aim is to understand the scale and scope of online hateful content, taking into account its different forms, from 'everyday' subtle actions to overt acts of aggression and criminality, and the different targets, such as ethnic minorities and women. The project also aims to understand the dynamics and drivers of hate, providing granular insight into when, where and why it manifests.

Helen  
Margetts

**Principal Investigator**

**Humanities &  
Social Sciences  
Faculty**



# Multi-scale Infrastructure Systems Analytics

---

The MISTRAL project is an EPSRC Programme carried out by the UK Infrastructure Transitions Research Consortium (ITRC) of seven universities and over 50 partners from infrastructure policy and practice. The project is developing an integrated view of the urban water cycle, coupled within the existing NISMOD national models of water supply and waste water treatment. For the first time, building-level models of whole cities are being used to represent the whole urban water system (waste water and stormwater). This will allow better analysis strategies for urban drainage, such as separation of waste water and stormwater systems, and their associated costs and performance.

Chris  
Kilsby

**Principal Investigator**

Science, Agriculture  
& Engineering  
**Faculty**

# Manuscript Print Digital

Ruth  
Connolly

Principal Investigator

Humanities &  
Social Sciences  
Faculty

The School of English Language & Linguistics runs group teaching projects using a 17<sup>th</sup> century cookbook manuscript, written by Jane Loraine, who lived in Northumberland. She is likely to have been the wife of Nicholas Loraine, son of Ambrose Loraine of Hartburn, and probably a member of the Fenwick family. The cookbook is firmly rooted in Northumberland. There are 67 recipes attributed to 41 individuals in this cookbook, only 13 of whom are men. RSEs are providing advice and support to maintain a digital archive of the texts that is added to by students on annual group projects.



# Organ Retrieval Decision Support

---

Organ retrieval and successful transplantation relies on accurate assessment of organ viability at the point of retrieval. Typically this is done by the retrieval team and relies on their combined knowledge and experience. This pilot project looks at how machine learning might provide decision support for surgeons in the field, taking a small number of images of pancreases labelled by expert transplant surgeons to produce a model that could reliably grade new images of pancreases as good or bad. The pilot will assess whether given more time and more images, a model with greater accuracy is achievable.

Colin  
Wilson

**Principal Investigator**

Medical  
Sciences  
Faculty

```

    //waits for the appear event when appropriate
    var check = function() {
        //is the element hidden?
        if (!t.is(':visible')) {
            //it became hidden
            t.appeared = false;
            return;
        }

        //is the element inside the visible window?
        var a = w.scrollLeft();
        var b = w.scrollTop();
        var o = t.offset();
        var x = o.left;
        var y = o.top;

        var ax = settings.accX;
        var ay = settings.accY;
        var th = t.height();
        var wh = w.height();
        var tw = t.width();
        var ww = w.width();

        if (y + th + ay >= b &&
            y <= b + wh + ay &&
            x + tw + ax >= a &&
            x <= a + ww + ax) {

            //trigger the custom event
            if (!t.appeared) t.trigger('appear', settings);
        } else {
            //it scrolled out of view
            t.appeared = false;
        }
    };

    //create a modified fn with some additional logic
    var modifiedFn = function() {
        //mark the element as visible
        t.appeared = true;
        //is this supposed to happen only once?
        if (settings.one) {
            //remove the check
            w.unbind('scroll', check);
            var i = $.inArray(check, $.fn.appear.checks);
            if (i >= 0) $.fn.appear.checks.splice(i, 1);
        }

        //trigger the original fn
        fn.apply(this, arguments);
    };
}

```

45        //trigger the original fn  
       fn.apply(this, arguments);  
 };        //the modified fn to the element  
 t.appearOne('appear', settings.data, modifiedFn);

# Research Impact Fund: HaSS

---

The Digital Institute was awarded a three year fund to support the HaSS faculty through the application of digital technologies, tools and practices. The RSE team has used this pot to fund a series of pilot project calls. This was decided as a way to reach as many researchers as possible to demonstrate the potential of technology in their research areas. The fund targets researchers who wish to use pilots or demonstrators to leverage further external funding. Pilots are run for three months at 0.5 FTE so that two projects can be run concurrently. The hope is that the grant applications that follow on from these pilots provide a long term funding stream to keep an ongoing RSE team financially viable.

Paul  
Watson  
**Principal Investigator**  
  
**Humanities &  
Social Sciences  
Faculty**



# Senescent Cell Counting

---

Research into developing therapies that have potentials to increase healthspan has an important implication for the future healthcare. Recent research indicates cell senescence as a primary driver for age-related disease, because pharmacological interventions that preferentially kill senescent cells (senolytics) have been found to be able to postpone, or cure, a wide range of pathologic conditions. To progress anti-senescence therapy, senolytics drug screening work is underway. The RSE team was approached to conduct some exploratory work on using machine learning to distinguish and count young (labelled red) and senescent cells (labelled green) in single fluorescent microscope images before and after the drug treatment. Such tool can accelerate the accuracy and speed of the discovery of novel and better senolytics drugs.

Thomas  
von Zglinicki

**Principal Investigator**

Medical  
Sciences  
Faculty



# Smart Cities & Inequality

---

This project aims to pool data streams from the Newcastle University Urban Observatory (UO) with ONS residential/workplace zone demographic data, as well as other fine-scale demographic information. Data will be used to estimate 'deserts' in sensor coverage and areas of high data uncertainty and to link with socioeconomic and demographic characteristics of day, evening, and transitory populations. The goal is to estimate sensor coverage of vulnerable populations (e.g. the elderly or poor), to develop measures of "population at risk" for use with sensor data products, and to increase understanding of links between dynamics of population movements and sensor coverage.

Rachel  
Franklin  
**Principal Investigator**

Humanities &  
Social Sciences  
**Faculty**

# Streaming Data Modelling

---



This project addresses one of the key challenges of the big data age: the development of scalable algorithms for extracting useful information from large, complex, heterogeneous and ever-growing data sets in (near) real-time. Integrated modelling of data sources from streaming data networks will typically require the inversion of explanatory stochastic models. Although significant computational and methodological advances have been made in this area in recent years, true simultaneous inference for both static parameters and dynamic states of non-trivial models remains challenging. New methodological innovations will be developed, motivated by stress-testing robust new implementations of the best available on-line inferential algorithms in two demanding application areas.

Darren  
Wilkinson

**Principal Investigator**

Science, Agriculture  
& Engineering  
**Faculty**

# OUTREACH



## HIGHLIGHTS

# Conferences & Workshops

---

It is important that RSEs make time to showcase their work and interact with the wider academic community. Submissions to conferences and workshops are actively encouraged.

## RSE CONFERENCE

The entire team attended the flagship RSE conference in Birmingham this year. Mike gave a talk on visualizing uncertainty.

## INNOVATION FESTIVAL

Andy joined the NICD team for a two-day hackathon hosted by Northumbrian Water.

## DIGITAL HUMANITIES

Mark and Fiona travelled to Utrecht in the Netherlands to attend the international Digital Humanities conference.

## N8 RSE

Kate and Jannetta attended the first N8 RSE workshop in Leeds to meet other RSEs from across the N8 network.



## HIGHLIGHTS

# Attending Training

---

Working in a technical specialist role it is important to keep skills up to date and explore new areas of knowledge. Team members are afforded time for both technical and soft skill personal development.

## RSE CONFERENCE WORKSHOPS

The third day of the RSE conference was given up to workshops. The whole team attended a variety of different sessions.

## ASPIRING LEADERS

Mike attended the RSE Aspiring Leaders workshop in Southampton to learn from skills from existing group leaders.

## REDHAT OPEN SOURCE WORKSHOP

A number of people from the team attended a workshop on open sourcing code hosted at the Alan Turing Institute.

## NU MENTORING

Several members of the team have attended training with the NU mentoring service as mentees or mentors.



## HIGHLIGHTS

# Delivering Training

---

The team has had a number of opportunities to pass on its knowledge and skills to others. The delivery of training will be an area of potential growth for us in 2020

## CODE COMMUNITY

Mike, with the help of a small team, runs a bi-monthly meetup group for researchers around the university who write code.

## SOFTWARE ENGINEERING

Mark gave a presentation in the School of Engineering to postdocs and researchers on the basics of how to manage a code base.

## VERSION CONTROL

Jannetta delivered a training course for the CESI project on how to use version control to manage code.



## HIGHLIGHTS

# Student Supervision

As well as delivering training to staff, the team is encouraged to offer student projects and get involved in teaching at both undergraduate and postgraduate levels.

## WEB DESIGN & UI/UX

Mark gives two annual guest lectures to undergraduate and MSc students on UI/UX first principles.

## GAMES MSc

Frances supervised a student to create a game that utilised some of the modelling work from the CANDO project.

## CONVERSION MSc

Mark supervised three students with their individual projects, mostly working on tools that help run the RSE team.

## NATIVE AMERICAN COLLECTION

Mark supervised an MSc student to deliver a website for the Great North Museum that showcased their Native American collection.

**HIGHLIGHTS**

# Volunteer Work

---

Several members of the team give up their free time for a range of worthy causes. Team members are actively encouraged to consider volunteering and are given flexibility to pursue opportunities.

**CODER DOJO**

Jannetta runs a Coder Dojo for local children every Saturday, based at the Rivergreen Centre in Durham.

**#TECHMUMS**

Kate helped organise and run a multi-week training course to help women get started in technology.

**WOMEN'S ENGINEERING SOCIETY**

Jannetta & Laura Brown were awarded £3000 to build an exhibition to showcase different disciplines of engineering.

**MAKING FOR THE LIGHT**

Frances was a panel member fielding public questions on the CANDO project and the pitfalls of developing medical technologies.

# TESTIMONIALS



**Stuart Barr**

Professor of Geospatial Systems  
Engineering

Researchers at Newcastle have developed NISMOD-DB++ a dedicated database management system for interdependent infrastructure systems modelling and simulation as part of the £5.5m EPSRC MISTRAL (Multi-scale Infrastructure Systems Analytics) programme grant. NISMOD-DB++ provides a highly functional API for programmable model access to the data and analytical tools developed within the MISTRAL project. However, until recently there was no easily accessible user interface for more general user interaction with the system. Working with the MISTRAL researchers in the Spatial Analytics and Modelling Lab (SAM) the Research Software Engineering team has developed a map driven user interface to NISMOD-DB++. Using open source spatial data libraries this allows users to undertake spatially driven knowledge discovery of the data-sets held within NISMOD-DB++ and undertake queries to map and retrieve data-sets. A particularly novel development provided by the Research Software Engineering team is the ability to develop linked queries across multiple databases held within the hybrid database architecture of NISMOD-DB++. This provides a unique capability to retrieve disparate but related data-sets according to logical, topological, spatial and temporal relationships. For example, it could retrieve details of the households that live in a detached property whose electricity feeder network is supplied by a particular 11kV/440V electricity sub-station. Such queries involve the dynamic linking of multiple databases and the use of analytics such as real-time graph searches to retrieve spatio-topological relationships. The ability to work with highly skilled research software engineers has significantly accelerated the development of the query user interface to NISMOD-DB++ at a critical stage of the project.



**Jennifer Richards**

Director of Newcastle University  
Humanities Research Institute  
(NUHRI)

Working with Mark and his team has been a real pleasure. At ATNU we appreciate their readiness to engage with our ideas, and share their knowledge. The feedback we have been given on our shared work when showing it at conferences has been outstanding: a testimony to its quality. Our relationship with the Digital Institute through ATNU is unique in the UK, and it is helping us to create a new relationship with the Wellcome Collection that could be our future Impact Case Study. By the same token, I appreciate the readiness of Mark's team to work with the Research Institutes in HaSS, NUHRI especially, showing colleagues how they can enhance their research, and ask new/different research questions, while helping me to create links between colleagues across NU.



**Stephanie Klein**

ePRaSE Project Manager  
Pharmacy Directorate  
The Newcastle upon Tyne  
Hospitals NHS Foundation Trust

The Newcastle upon Tyne Hospitals NHS FT Pharmacy staff and academics from the Newcastle University School of Pharmacy have been working in partnership with Research Software Engineers from the Digital Institute since June 2018. They have helped deliver, in a tight timescale, the first iteration of a Web based tool being used nationally across the NHS for users to assess the maturity of their organisation's electronic prescribing systems. This project is part of the national Patient Safety and Medicines Optimisation Global Digital Exemplar programmes and its importance has already been recognised by the Healthcare Safety Investigation Branch. It has also raised the profile of Newcastle University in respect to both the Pharmacy School and capability of the Digital Institute, in addition to supporting the PhD of one of the lecturers. The research engineers' knowledge, skills and support will continue to be an important part of this project, which is funded to develop further iterations on an annual basis into 2022.

# CHALLENGES

## 1. FINANCE SUPPORT

---

Each faculty has its own finance support structure. This presents challenges when requesting help or making changes to SAP records.

## 2. GROWTH

---

Demand for services is larger than the resources we have available.

## 3. RETENTION

---

Retaining the skilled people we have is a priority for 2020 & 2021.

## 4. MARKETING

---

Improving the visibility of the team and its activities remains a challenge. There are long waiting lists when requesting updates to websites.

# 1. Finance Support

---

The largest procedural challenge facing us is the configuration of finance support across the University. Each faculty has its own structure and procedures for different types of finance requests. Typically this is fine, because an individual research group in a school will often have a single point of contact. However, when working across the University in the way that we do, this model breaks down because we then have a single point of contact per project. As of December 2019 we have 36 ongoing or proposed projects, meaning keeping a directory of names to contact when working on each project. There are time and opportunity costs and risks with working on complex project setups with so many different people. Financial reports are used when forecasting future resourcing needs, so it currently involves a lot individual requests to collate all the necessary information, taking up a high proportion of team management time.

Beyond reporting on a per project basis, it is especially difficult to arrange changes to SAP records when an RSE either changes project or changes their FTE split. Frequently there are situations where an RSE might be deployed on two projects in two different faculties. To give the required oversight to the different finance teams it ends up being the RSE team coordinating the changes to SAP (usually via email) by connecting two business partners who may never have spoken to each other before. Even before starting on a project, generating costings to be able to give PIs quotes for our time on a project is a challenge. Depending on the school or the faculty, we have to find the finance partner for the project RES account and ask them to generate a costing on our behalf.

## 2. Growth

---

The rapid growth of the team is causing all the same pains that tech startups go through in their initial phases; the ad-hoc processes, fractured decision making and difficulties forecasting the requirements for future potential projects. Tied into these challenges is the idea of reputational cost. Due to being a new entity we have no reputation around the University, either good or bad. In the first couple of years we need to say "yes" to as many projects as possible, otherwise we risk becoming the group that says "no" – and researchers will not seek us out in the future. In order to meet these demands we have appointed multiple people in three different recruitment rounds. This cannot be a long-term solution though, eventually there will be a plateau in the number of projects we need to maintain staff levels. The primary aim when it comes to recruitment is to make sure that good, highly skilled people continue to have a role here.

At current staff levels the management of the team is at its limit of what can reasonably be done to maintain the function of the team and the happiness of the people within it. Staff development activities such as one-to-ones and the PDR process are essential to the functioning of the team but take up more and more management time as the team grows. At some point there will be the need for extra management support to these activities. This will free up more leadership time to develop processes, pursue strategic aims and continue to bring in more projects to fund the long-term viability of the team.

## 3. Retention

---

From the first 12 months of experience, it's clear there is enough demand for our services from across the University and its partners. If anything, there is too much demand, and this is causing problems trying to smooth the flow of incoming work. Ideally, we wouldn't take on any more work until spring 2021 when the first contracts come to an end. As discussed in the previous section, because we're a new entity we are not yet able to turn down work due to the reputational cost. Continually recruiting is not a sustainable long-term strategy as it harms the prospects of staff retention in leaner times. In order to meet our project needs in 2020 it may be that we look at alternative ways of bolstering our resource through the use of different recruitment methods such as internships or secondments.

As seen in the Team section, we've been exceptionally fortunate with our recruitment in 2019. We've been able to put together a team diverse in technical skills, knowledge of programming languages and career experiences. All of these individuals are employed on two year fixed-term contracts, so the focus for 2020 and 2021 must be to secure enough funding to retain everybody in the team. Around 40% of our projects come at the post-award stage, which is much harder to plan for. In those cases researchers have tight deadlines and give short notice of need. Based on experiences this year there is no doubt that there is enough work to sustain the team, the challenge will be getting the end of one project and the start of the next to line up.

## 4. Marketing

---

In order for the team to meet its primary objective of supporting the development of software in research, researchers need to know we exist and how to request our services. Throughout 2019 we have relied largely on existing relationships, a few people in key strategic positions and word of mouth. Whilst all of these methods of finding work have their place, more effort is needed in marketing ourselves around the University. The two focus areas will be the creation of pages on the Newcastle Data website and a speaking and engagement tour across the University.

The creation of sub pages on the Newcastle Data website is an urgent task for 2020. Expertise exists to do it within the team, but these people are fully allocated onto research projects. Content for the website can be created locally, we have access to text and images and can write more as needed, but we need more support from marketing

to create and maintain the website. In the meantime it is possible to create intranet pages on SharePoint so that researchers can access some basic information about our services.

Speaking around the University is an easier problem to solve. Schools across all faculties usually have seminar series or general meetings that can be used to request 10-minute slots rather than one-off events that can be poorly attended. These types of events are useful for reaching researchers, but it is also useful to speak with other research support staff such as research managers. Again, these individuals often have team meetings that can be used to request speaking slots. The issue with all of these is simply identifying who the right people to speak to are in order to arrange time in an existing agenda.

# OPPORTUNITIES

## 1. TRAINING

---

The team is highly skilled and is an underused asset when it comes to the training of other staff around the University.

## 2. TEACHING

---

The team can offer internship style MSc student projects.

## 3. The Turing

---

The team can contribute to capability and capacity to the Turing RSE network

## 4. UKRI e-INFRASTRUCTURE

---

Recent changes in UKRI have seen one of the founders of the RSE movement, James Hetherington, installed as director of the e-Infrastructure program.

# 1. Training

---

Taken as a whole, the University produces a vast amount of code. This encompasses everything from small scripts that help individuals perform menial tasks, code that underpins key processes within research groups and all the way up to large scale software deployments in flagship projects such as the Urban Observatory. Typically students and early career researchers get some help on how to write code, but little to no training on how to manage and maintain it. This problem is not unique to Newcastle and is merely one symptom of the larger crisis of reproducibility in science. If a piece of code is written to analyse data, and that data (or the graphs derived from it) are part of a publication then the code should be stored and maintained. Without it, papers cannot be reproduced.

The tools and techniques used to manage code are the same whether it is a single small script or a large software deployment. As a team we give talks and encourage researchers to use these tools and learn to do things the right way at the start of their work to save pain further down the line. We currently do not have the time or teaching skills to provide much more than this, however this is a hugely underused resource. With proper instructor or teacher training, the skills the team possesses can be passed on in a more strategic, workshop-based way – improving the basic code management skills of every researcher, thereby ensuring academic outputs from Newcastle that rely on code are reproducible. Ideally, we should pursue membership of Software Carpentry [D] to gain access to their training courses.

## 2. Teaching

---

Over the last couple of years the Digital Institute has offered a small number of MSc projects to computer science students, many with great success. In most cases these MSc projects have at least been alongside existing research projects – if not entirely within them. Sometimes students can contribute so much that we seek to employ them on completion of their degrees. The quality of these assignments has been lucky rather than strategic, projects are offered on a first come first served basis, which does not always lead to the best outcome for either the project or the student. Being able to offer projects, but then choose students based on grades or some other informal interview process would be better. The projects we offer are current real-world problems, and to continue to offer projects at all there needs to be some expectation that the student assigned to it can deliver.

In the longer term, another way to improve the team's offer to students could be the creation of an internship programme where these types of student projects become more formal. Students would be expected to be physically located with the team, directly contributing to projects and being part of the wider group. This model would still require students to submit a dissertation length summary of their experiences and they would be encouraged to view the position as work experience that would be useful experience for their future employment. Such a programme might also open the door to students from outside of solely computing science, providing they had a sufficient level of coding experience.

## 3. Alan Turing Institute

---

The Alan Turing Institute's (The Turing) Research Engineering Group (REG) [E] is in the same position as us, massively in demand and unable to recruit fast enough to fill that demand. One model that the Turing has championed and is trying to kick start is an RSE network where any of the partner institutions can request resource or make resource available. The primary aim of the network is to share knowledge, expertise and best practices. This ongoing exercise in capability and capacity building means that for Newcastle to see benefit from the network we need to contribute to it. We'll get more out if we play an active role and contribute ideas, skills and resource.

The network has already been beneficial to us as there are several extra projects that we've picked up that are additional to the main projects that come with the Turing Fellows. Hate Speech and Learning Machines are the main examples. If, as expected, the Turing continues

to experience high levels of demand then Newcastle is well positioned to leverage our existing connections to make the most of any new opportunities.

There are some issues that are still being worked through that would enable a network to function efficiently. Chiefly the Master Services Agreement (MSA) which governs the specifics of how Newcastle collaborates with the Turing is going through several rounds of review. Until then we need separate agreements per project. The driver behind the MSA is to make the contractual process of exchanging work with the Turing network less burdensome. Ideally this will be resolved in the second half of 2020 so that we can pitch for more collaborative projects with Turing partners and strengthen our connections with fellow RSE teams.

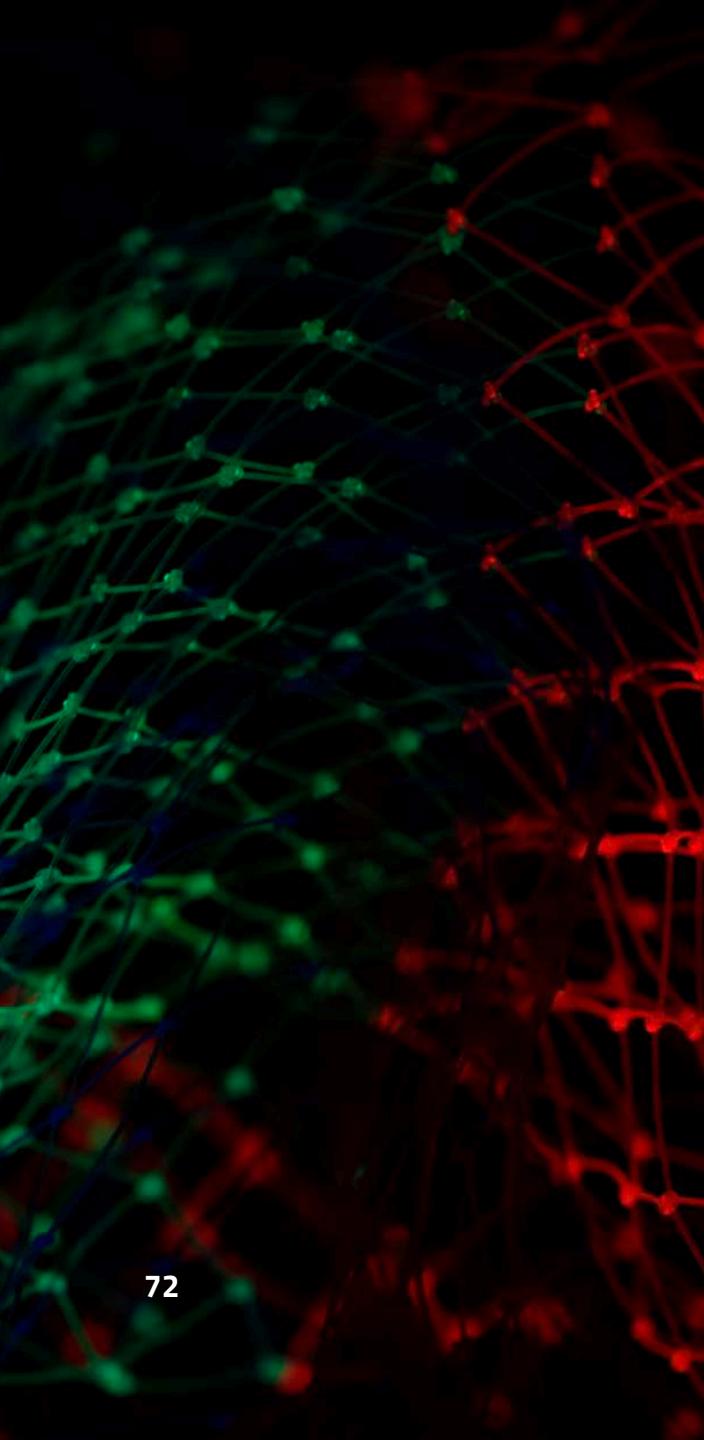
## 4. UKRI e-Infrastructure

---

When UK Research and Innovation (UKRI) was created in 2018, one of its aims was to meet the ambitions of the Government's Industrial Strategy and the target to increase total R&D investment to 2.4% of GDP by 2027. In order to do this the UK will need to invest in research and innovation infrastructure. On behalf of the Department of Business, Energy and Industrial Strategy, UKRI sought to understand the UK's current capabilities and to plan for the future, by conducting an assessment of the UK's research infrastructure for the first time. This report [F] assesses the future infrastructure landscape and identifies needs, opportunities and key themes that could be a major benefit to the UK's capability to 2030. It is intended as a strategic guide to inform investment decisions for the next generation of infrastructure.

RSEs are mentioned throughout the report as being a particular subset of the infrastructure landscape that UKRI is looking to invest in. Indeed, Theme 6: Software and skills, mentions career development for RSEs as critical for the UK to maintain its global standing in research. When looking to the future, UKRI is aware of just how much research depends on software; "The effective operation of any e-infrastructure ecosystem is also dependent on the availability of the right skills and talent in universities, research institutes and business. In this sector, support for research software engineers and research data professionals is particularly critical" [F]. UKRI will be looking to both fund grants that have solid software plans and RSE activities more directly. One of the founding members of the RSE movement in the UK, James Hetherington, left his post at the Alan Turing Institute in December 2019 to take up a new role as Director of e-Infrastructure at UKRI.

# 2020 PROJECTS



# Auto Generation of Optimal Deep Learning Networks

---

Deep Learning has the potential to solve many of our Big Data challenges. This has been demonstrated through prior work in such areas as image classification and voice to text. However, there is currently no known approach to determine the 'best' Deep Learning network for solving a particular problem. Where best is normally defined to be the network with highest accuracy, with developers often just constructing, and testing, different networks until they reach a level of accuracy which is acceptable. The ability to derive better Deep Learning networks would have far-reaching and significant impact across both industry and academia. This project will make inroads into this problem domain, seek better understanding of the research challenges and develop an application to an external funding organisation.

Steve  
McGough

**Principal Investigator**

Science, Agriculture  
& Engineering  
**Faculty**

# Climate Model Data Cleaning

---

Researchers in the School of Engineering need to develop scripts to verify files containing climate model data. The verification is non-trivial due to the sheer quantity of files and their size on disk. The total expected download will be in the region of 150TB and each file needs to be checked for repeated or missing data, correct time periods and implausible values. The RSE will deliver a suite of scripts that iterate over the files and can be used against future data releases.

Selma  
Guerreiro

**Principal Investigator**

Science, Agriculture  
& Engineering  
**Faculty**



# Do Machines Have a Sense of Place?

---

The RSE team has been asked to do an exploratory piece of work on how machines understand a heritage location based on image recognition. Different algorithms can be used to pick out what is depicted in an image. However, each algorithm prioritises different dimensions of the image, such as its colour or the objects in the scene. How does this difference affect a machine's understanding of a place such as a town square? And, can machine vision contribute to a collective understanding of a place's meaning?

Areti  
Galani

**Principal Investigator**

**Humanities &  
Social Sciences  
Faculty**



# EpiChange: Quantifying Changes After Epilepsy Surgery

---

Resective surgery for epilepsy, where the part of the brain thought to cause seizures is removed, leads to seizure freedom in around 70% of patients one year post-surgery. This falls to around 50% at five years post-surgery. It is not fully understood why surgery only works initially for some patients, and why this falls over time post-operatively. Using univariate and multivariate data analysis, in conjunction with machine learning, we will learn how brain dynamics change after surgery, and if this change relates to outcome. Crucially, we will attempt to identify which factors in brain dynamics correlate with seizure relapse, even years after surgery.

Peter  
Taylor

**Principal Investigator**

Science, Agriculture  
& Engineering  
**Faculty**



## FinTrust: Trust Engineering for the Financial Industry

---

The FinTech industry is one of the major growth industries in the United Kingdom. These companies create new, cheaper and faster services, utilising the latest technologies such as cloud, mobile and blockchain. To succeed they need to gain the trust of customers in a period that society's trust in the financial industry is still impacted by the mortgage crisis almost a decade ago. They need to gain this trust while technologies are changing rapidly and data breaches are continuously in the news. FinTrust will research the issue of trust in FinTech, identifying the generic research challenges and establishing fundamental research results. A particular focus will be on increased automation through the use of machine learning algorithms, which may have implications that affect consumer trust in the new services.

Aad  
van Morsel

**Principal Investigator**

Science, Agriculture  
& Engineering  
**Faculty**

£ 41,574



# Human Cell Atlas

---

Cells are the fundamental units of life, and the human body contains around 37 trillion of them. Most cells contain the same genome, but gene activity varies from cell to cell. To truly understand the genome, we need to understand how it instructs cells to carry out their unique functions in the body. A new global initiative called the Human Cell Atlas is setting out to tackle this challenge, using powerful genomics approaches to define the cell types in the human body and reveal how they behave in health and disease.

Muzlifah  
Haniffa

**Principal Investigator**

Medical  
Sciences  
Faculty



# Learning Machines

---

Machine learning techniques are effective for building predictive models because they are good at identifying patterns in large datasets. However, the development of a model for complex real life problems often stops at the point of publication or proof of concept. This project addresses the issue of keeping models updated, in order to reflect current environment.

In the medical and criminal justice domain, data collecting and labelling under real world circumstances specifically for the application of machine learning is rare. The ability to store data to generate newly labelled data enables models to be updated when data changes to reflect trends over time. This project will develop a generalised infrastructure for versioning new labelled data, so that they can be used to train and retrain models, and keep them updated with changes over time.

The research question posed by the Learning Machine project is: What are the common ways in which newly labelled data can change over time? How do we automate detection of changes, so that models can be retrained, and therefore kept updated to reflect current environment?

Alan  
Wilson

**Principal Investigator**

Science, Agriculture  
& Engineering  
**Faculty**

# Newcastle University LAB Digital Edition Publishing Cooperative

James  
Cummings

**Principal Investigator**

Humanities &  
Social Sciences  
**Faculty**

The Newcastle University LAB Digital Edition Publishing Cooperative Implementation and Containerisation Project (NU-LAB) is a sub-project which will test the reusability of the main project outputs and determine barriers to re-use by organisations external to the main project. It will do this by creating a test installation of the LAB|Pro software specifically geared for the creation of scholarly digital pilot projects. It will implement several case studies in the installation that test a variety of features. In doing so, where there are features in need of improvement, this work package will collaborate with the project in creating them as generalised features.



## Quality in Organ Donation

The Quality in Organ Donation (QUOD) Biobank was established in 2012. This unique resource combines collection of detailed clinical information from virtually all organ donors in the UK with blood and urine samples taken around the time of donation and carefully collected small biopsies from a range of organs stored within a central 'bank'. This has been invaluable in research focused on understanding how stress associated with becoming an organ donor around the time of death affects control of important whole body systems such as blood pressure and glucose levels in addition to impact on specific organs. This has already enabled otherwise impossible research focused on better selection and optimisation of organs, enhancing successful transplantation.

James  
Shaw

**Principal Investigator**

**Medical  
Sciences  
Faculty**



# Rome Transformed

---

This project will pioneer a radically new methodology designed to analyse complex urban landscapes, exploring buildings buried up to 10 metres below the modern ground surface. Its focusses on a 'forgotten' quarter of Rome which, while omitted from most tourist itineraries, served as home to emperors and popes for generations. Between the first and eighth centuries AD, many of the most powerful people on earth lived in and around the Caelian Hill in the south-east of the city. Drawing together diverse strands of data to visualise the way this area changed over eight centuries, the team will examine in detail the character of its many features, from palaces and the world's first cathedral, to fortifications, aqueducts and private homes. This will reveal in turn how these features related to each other and to prevailing political, military and religious ideas.

Ian  
Haynes  
**Principal Investigator**

**Humanities &  
Social Sciences  
Faculty**

# APPENDICES

# Acronyms

**SAgE** | SCIENCE,  
AGRICULTURE &  
ENGINEERING

**HaSS** | HUMANITIES &  
SOCIAL  
SCIENCES

**FMS** | FACULTY OF  
MEDICAL  
SCIENCES

**DI** | DIGITAL  
INSTITUTE

**NICD** | NATIONAL  
INNOVATION  
CENTRE FOR DATA

**RSE** | RESEARCH  
SOFTWARE  
ENGINEER

# References

- A**   |   **RSE: STATE OF THE NATION**  
<https://bit.ly/rse-nation>
- B**   |   **STRATEGIC PRIORITIES FUND**  
<https://bit.ly/ati-spf>
- C**   |   **RSE Team Project List & Financials**  
Available on request from  
[mark.turner@ncl.ac.uk](mailto:mark.turner@ncl.ac.uk)
- D**   |   **SOFTWARE CARPENTRY**  
<https://bit.ly/s-carpentry>
- E**   |   **ALAN TURING INSTITUTE REG**  
<https://bit.ly/ati-reg>
- F**   |   **eINFRASTRUCTURE REPORT - UKRI**  
<https://bit.ly/ukri-ei>