



# Research Software Engineering 2020



# Contents



## 3. Overview

Introduction to the report, operating model and review of 2020.

## 31. Projects

Overviews of all the projects worked on during 2020.

## 76. Testimonials

Experiences of working with the team from academics.

## 12. Finances

Project analysis and funding breakdown by faculty, type and funder.

## 65. MSc Projects

Project contributions by our MSc Students.

## 80. Challenges

Looking ahead to some of the challenges the team will face in 2021.

## 19. The Team

Biographies of all the team members and their skill sets.

## 73. COVID-19

Contributions to the University's response to the COVID-19 pandemic.

## 85. Opportunities

Areas we can contribute more in 2021.

# Overview

Introduction to the report,  
operating model  
and review of 2020



# Introduction

2020 has been an exceptionally challenging year for everyone, COVID-19 has affected project priorities, UK research funding, access to the tools we need for our work and most importantly our ability to collaborate and grow as a team. The pandemic has brought the way models and data derived from research software are used. Government policy and our collective response to the pandemic now depends on it.

At the end of last year's report, I highlighted the challenges 2020 would bring and the need to balance staff retention with managing demand for access to the team. Given the other issues 2020 presented, those challenges haven't gone away, but we have taken some steps towards addressing them. Despite the challenging circumstances, the team has seen continued growth both in staff and project partnerships.

Most pleasing is that we're now known to a larger number of Newcastle academics and, crucially, project support staff, which enables us to get involved in grants much earlier. This allows work to be costed more

accurately and for our knowledge and skills help to inform the academic's ideas earlier in the process. The team has continued to expand its skills and knowledge, moving into the delivery of training and looking ahead to the skills needed to deliver future projects. The move to a new facility model will set us up to become more efficient at responding to researchers needs and simplify our project tracking and financial processes. In 2020, the team had a project income of £1.28m from 32 different projects, with a total research value of £42.96m to the University.

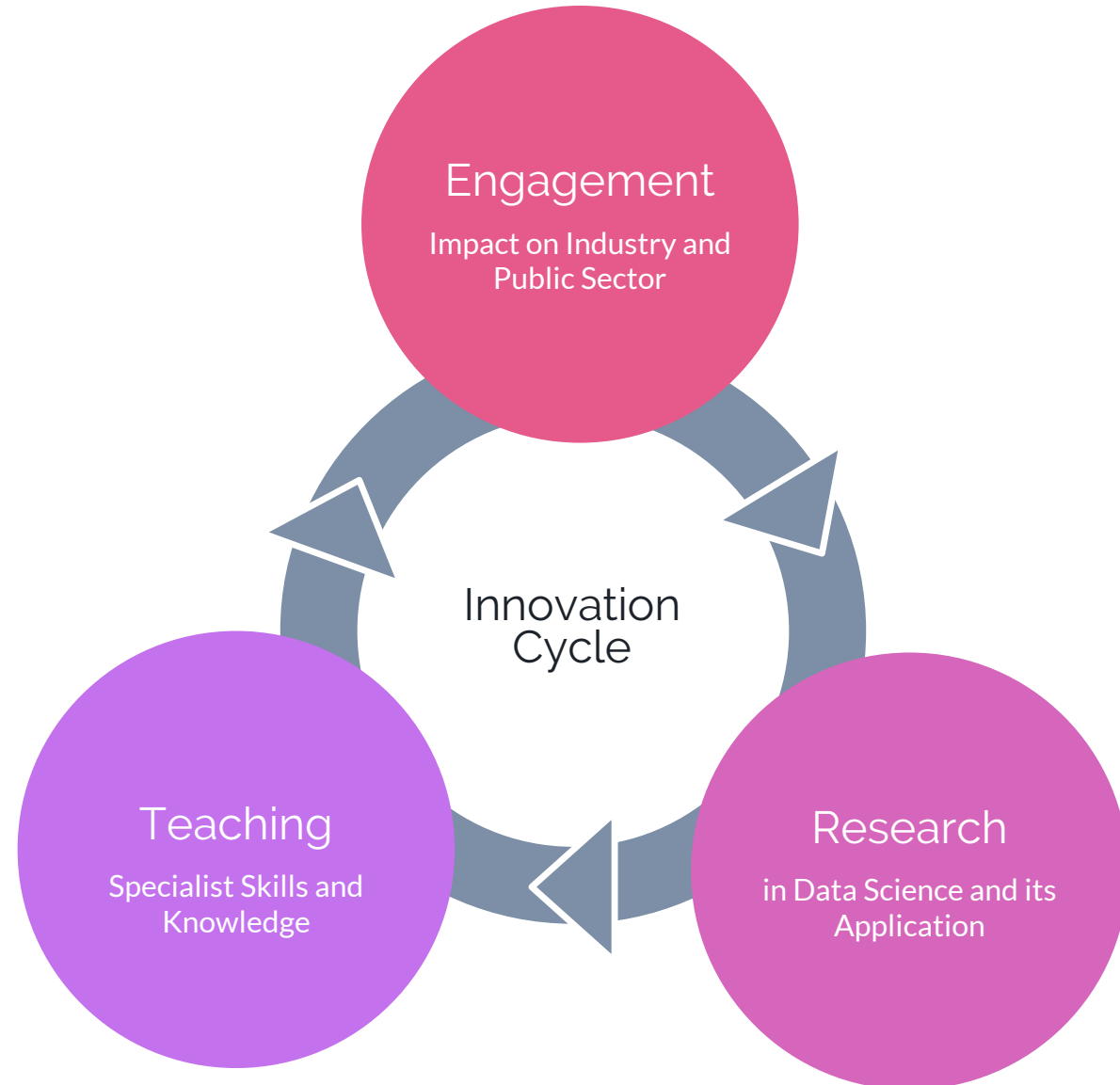
Mark Turner  
Head of Research Software Engineering  
Newcastle Data  
Newcastle University

[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.



# How We Work

## Combined Expertise

Costing RSEs onto grants gives you access to an entire team of experienced software engineers



## Fractional FTE Splits

RSEs can take on projects at less than 100%, the types of posts that are traditionally very hard to recruit for



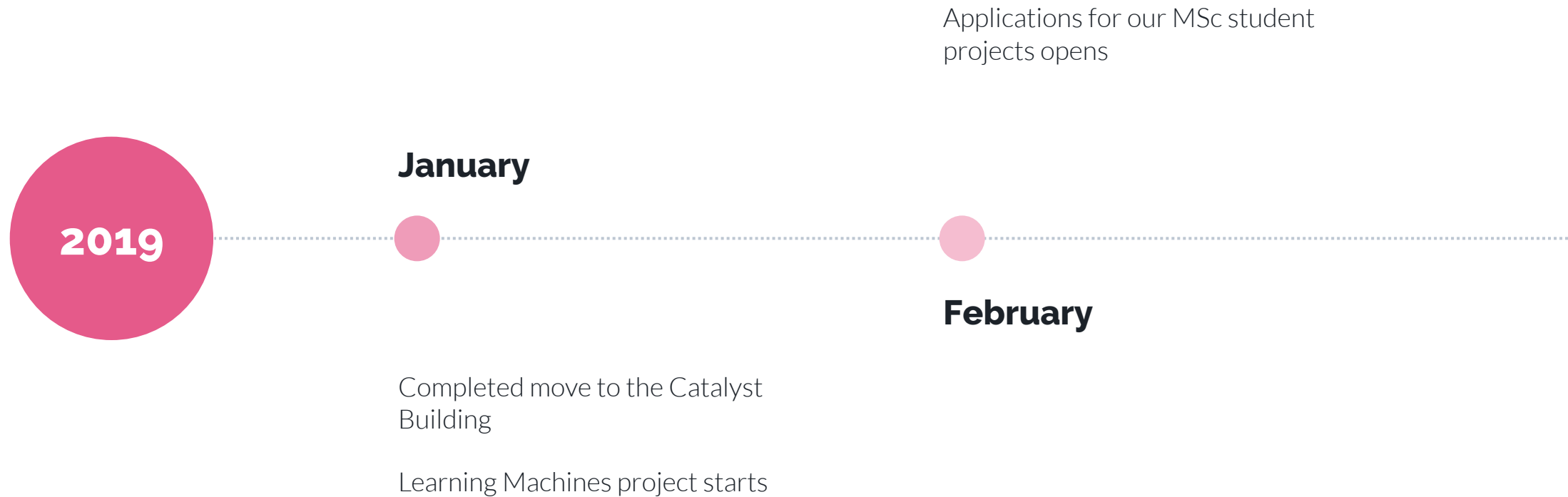
## Flexible Allocations

We can staff projects in intervals or with rising and falling resource needs





# Review of 2020



## March



Pro-Bono work on COVID  
projects

3<sup>rd</sup> RIF call opens

Jim McGrath & Dave Horsfall  
join the team

Rome Transformed & QuOD  
projects start



## April

## May



MSc students begin their  
projects



**June**



howbusyistoon.com launches

RIF calls 5 and 6 open

PlaSMA project starts



**July**

**August**



MSc students submit  
dissertations

**September**



The entire team qualifies as  
Software Carpentry Instructors

**October**



First full set of Software  
Carpentry workshops completed

**November**



UEDF bid successful



Nik Khadijah Nik Aznan joins the team

N8 HPC machine 'Bede' comes online

SWEET project starts

## **The Year Ahead**



**December**



**2021**

2021 is shaping up to be busier than ever. We've been involved in lots of grant proposals that will underpin our longevity and growth.

# Finances

Project analysis and  
funding breakdown  
by faculty, type  
and funder



# Overview

As forecast in the 2019 report, we have seen an increase in the projects coming from the Faculty of Medical Sciences. This has led to a shift in income share between the three faculties. The income share from FMS has risen from 11% to 25%, in turn reducing the SAgE share from 60% to 46%. HaSS income share remained the same at 29%.

## Medical Sciences

8 £322,073

Projects Income

25% £13,834,101

of Income Research Value

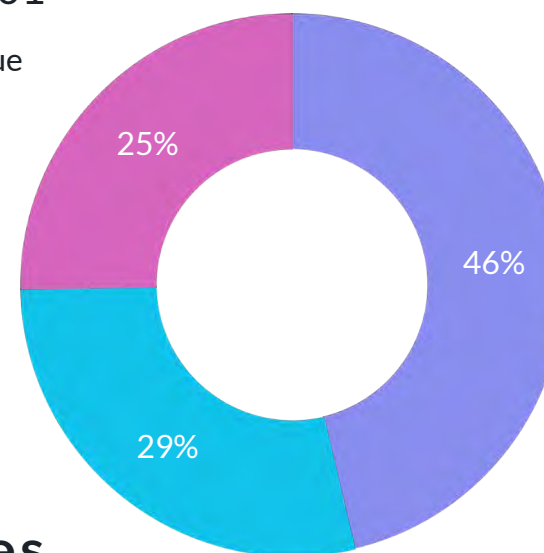
## Humanities & Social Sciences

8 £363,145

Projects Income

29% £3,189,776

of Income Research Value



## Science, Agriculture & Engineering

16 £590,643

Projects Income

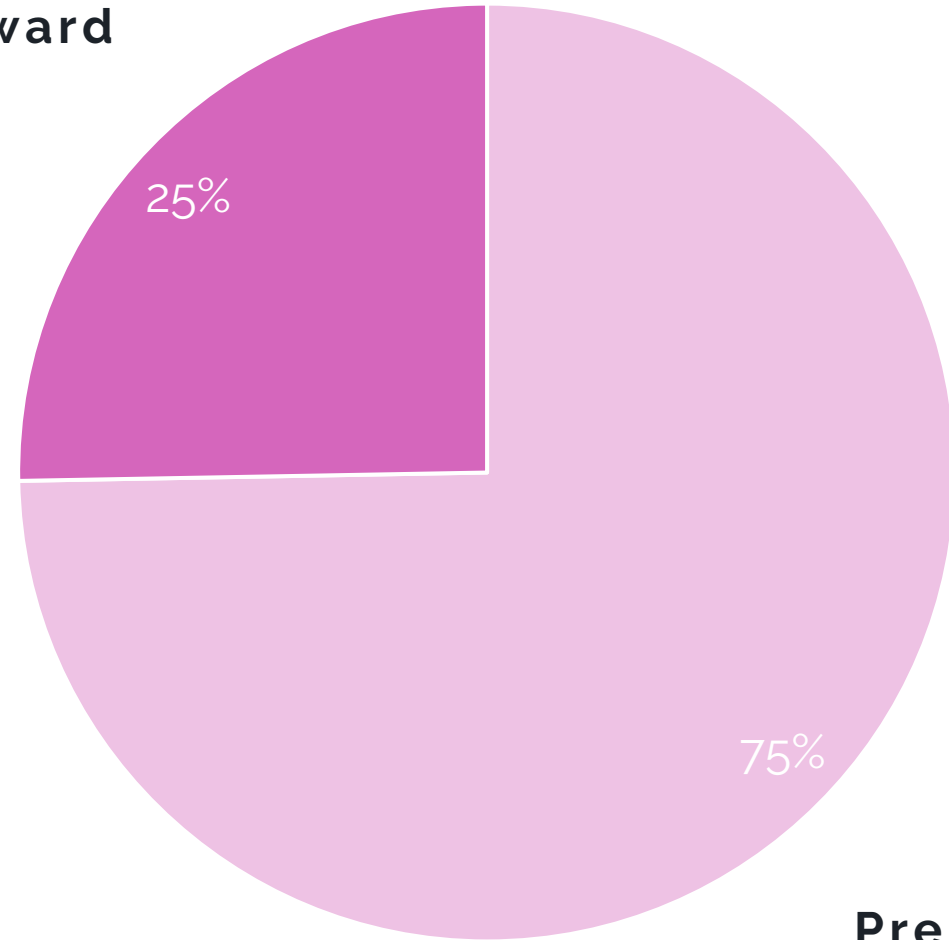
46% £25,929,534

of Income Research Value

# Award Stage

Work coming from projects already awarded by funders has grown since last year, up to 25% from 14%. This is likely caused by an increase in word-of-mouth leads to researchers with existing grants. We have also been involved in more grant applications in 2020, so the share of post award work may fall back again in future.

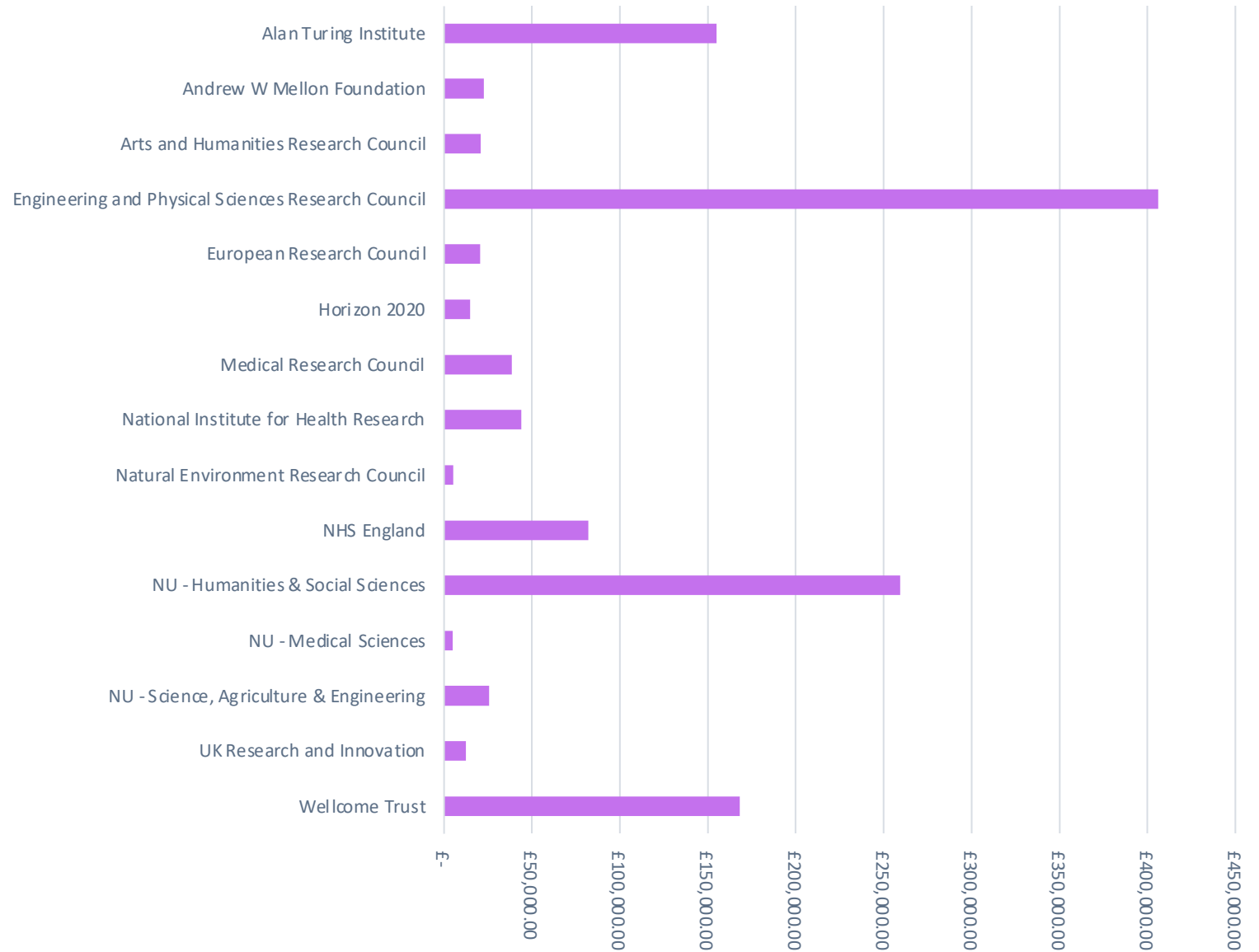
**Post-Award**



**Pre-Award**

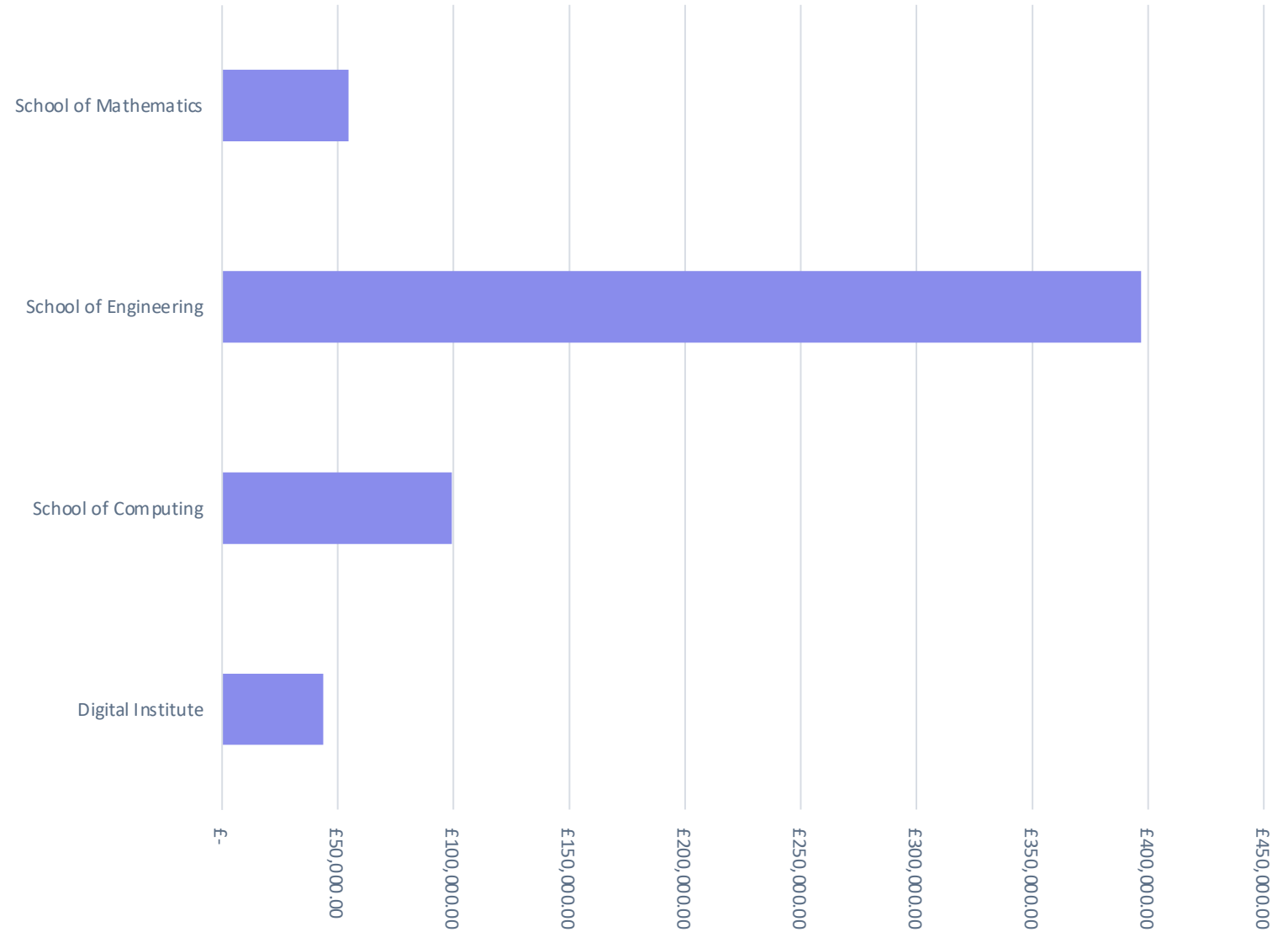
# Funder

Funder split remains as varied as previously, with increases in income from medical based funders reflecting the increase in work in those areas.



# SAgE Income

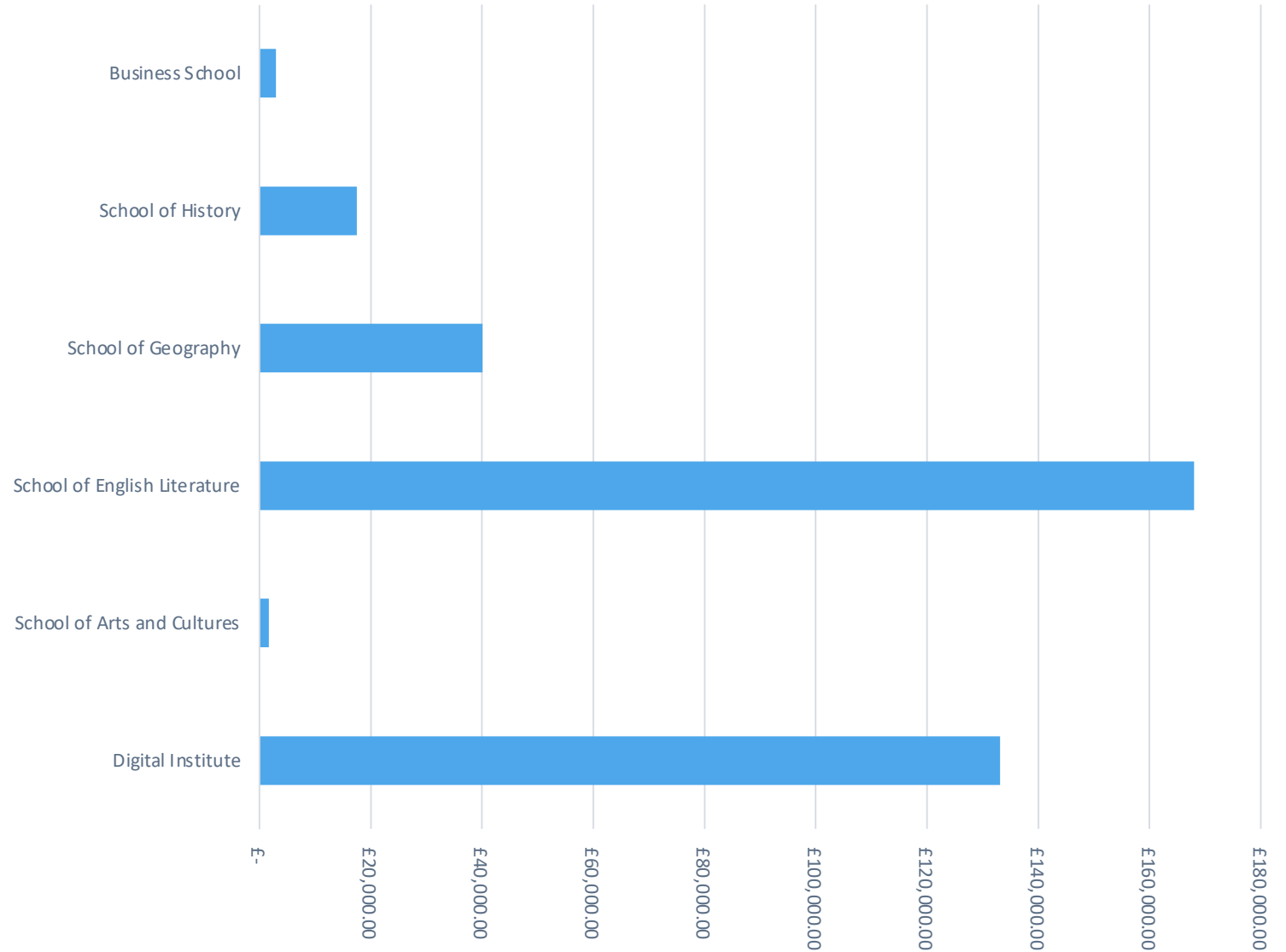
46% of our income comes from the Science, Agriculture & Engineering faculty. The majority of the income originates from the School of Engineering, the largest school in the faculty.





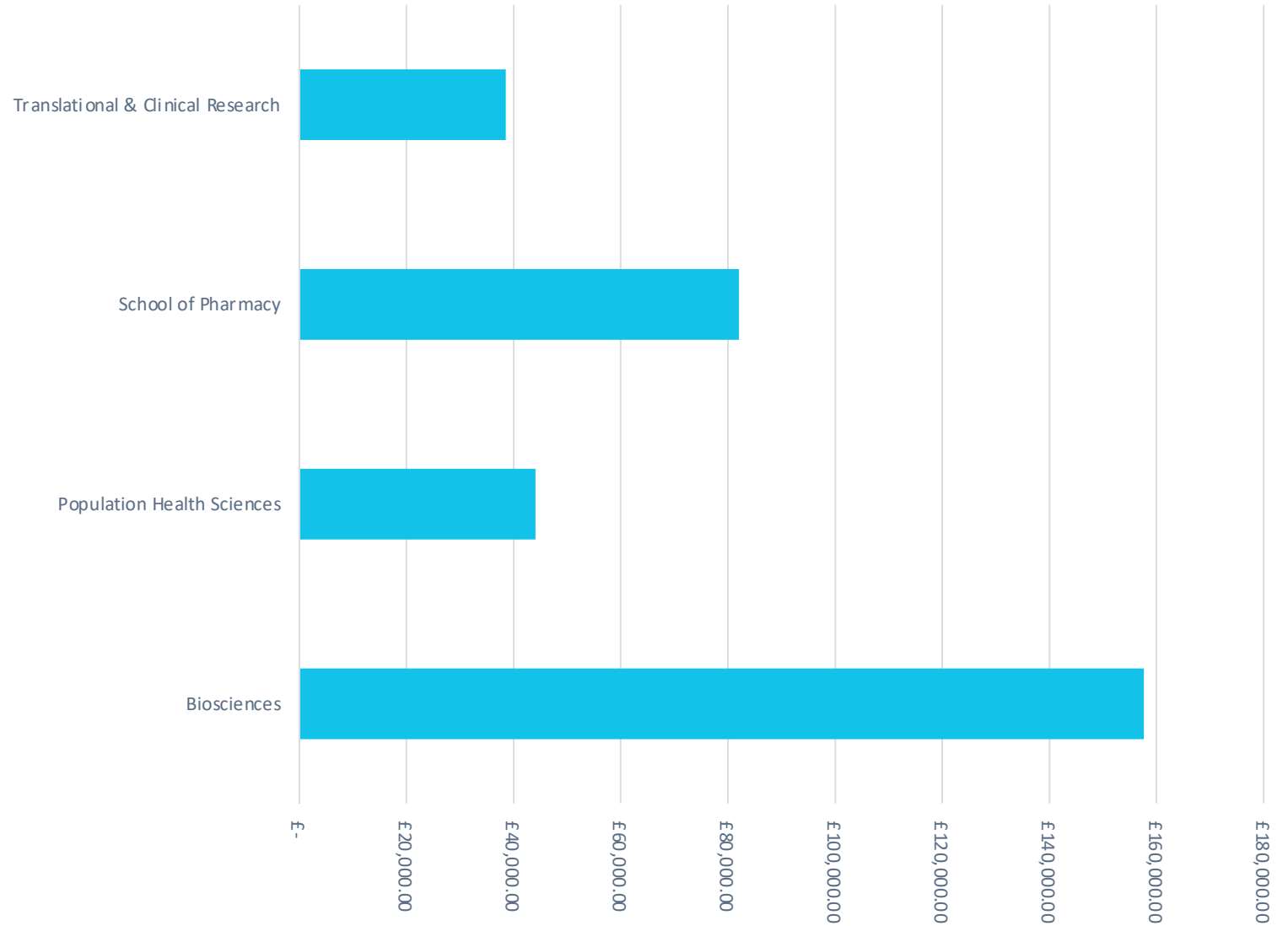
# HaSS Income

The faculty of Humanities and Social Sciences continues to be home to some of our most important strategic projects. The ATNU project with English Literature is entering it's final year and we will be looking to extend those relationships into the future.



# FMS Income

The Faculty of Medical Sciences was the largest growth area in 2020. We're pleased to have made some progress in our aim to be more visible in that faculty, despite COVID making that more difficult, and this has translated into more projects from across the faculty.



# Team

Biographies of all the team  
members and their skill  
sets

# Kate Court

## Research Software Engineer

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

## Research Software Engineer

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.



# David Herbert

**Research Software Engineer**

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.

# Dave Horsfall

## Research Software Engineer

Dave joined the team in March 2020, and works on the Human Cell Atlas and FinTrust projects. He graduated in Theoretical Physics from the University of Durham and has worked in both the open source community and the private sector. Dave is a full stack developer, with end-to-end development and deployment knowledge in public cloud platforms. He works mainly with Python and Laravel.





# Frances Hutchings

**Research Software Engineer**

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.



# Jim McGrath

## Research Software Engineer

Jim is an experienced RSE, having spent 4 years in the role at the University of Manchester prior to joining the team. He has worked in development and support roles with both public and private sector organisations either side of completing an MSc in Computer Science at Birmingham University. Jim is primarily a web developer with experience of both python and .NET frameworks, but he also has extensive knowledge of MS Office automation and programming.



# **Nik Khadijah Nik Aznan**

**Research Software Engineer**

Nik joined the team from Durham University where she was a PhD student focusing upon a multidisciplinary study combining the fields of Machine Learning and Brain-Computer Interfaces. Prior to that, Nik worked in the aviation and home-automation industries, as well as getting a Masters from South Korea. Nik brings experiencing of deploying machine learning systems using real world data, as well as combining complex models across data domains.



# Rebecca Osselton

## Research Software Engineer

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 developer in a range of languages and frameworks. She has an interest in supporting young people and women to engage in the technology sector.



A portrait of Mike Simpson, a man with short brown hair and a light beard, smiling. He is wearing a dark blue button-down shirt over a black t-shirt. The background is a blurred green and blue pattern. A large, curved magenta graphic element separates the portrait from the text on the right.

# Mike Simpson

## Research Software Engineer

Mike completed a BSc and MSc at Newcastle University, then worked as a Game Programmer before returning to study for a PhD. He is an adaptive programmer with years of experience in a range of languages, as well as an interest in interactive data visualisations. He has collaborated on a wide range of academic projects in his time at the university and has presented his work at a number of conferences around the world.



# Jannetta Steyn

## Research Software Engineer

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 computational neuroscience, 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

## **Head of Research Software Engineering**

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.



# Projects

Overviews of all the  
projects worked on  
during 2020





# Animating Text Newcastle University

Jenny Richards  
Principal Investigator

Humanities & Social  
Sciences  
Faculty


## Overview

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.

## Links

<https://research.ncl.ac.uk/atnu/>

<https://warden.atnu.ncl.ac.uk>



# Controlling Abnormal Network Dynamics using Optogenetics

**Andrew Jackson**

Principal Investigator

**Medical Sciences**

Faculty

## Overview

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.



# Centre for Energy Systems Integration (CESI)

**Laura Brown**

Principal Investigator

**Science, Agriculture and  
Engineering**

Faculty

## Overview

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.





# Climate Modelling Data Cleaning

**Selma Guerreiro**

Principal Investigator

**Science, Agriculture and  
Engineering**

Faculty

## Overview

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.



# Degree Apprenticeship Teaching Support

**Steve Riddle**

Director of Education  
School of Computing


**Science, Agriculture and  
Engineering**

Faculty

## Overview

Due to the COVID Situation, it was necessary to move most of the university's teaching material online for this academic year and with resources limited, the RSE Team offered to help support this transition. Mike Simpson worked with the School of Computing and the Institute of Coding in the run-up to Semester 1 to prepare and deliver online teaching material for the Degree Apprenticeship in Software Engineering, acting as module leader on the Introduction to Software Development module. This course is taught to part-time apprenticeships coming in from industry to prepare them for a career in software development. Material was prepared, recorded and then delivered via Canvas, the University's new online Virtual Learning Environment. Support for this course is set to continue into Semester 2.





# Do Machines Have a Sense of Place?

agora: 0.737

marketplace: 0.737

retail store: 0.737

building: 0.738

**Areti Galani**

Principal Investigator

**Humanities & Social  
Sciences**  
Faculty

greenishness color: 0.925

## Overview

This interdisciplinary exploration aims, on one hand, to provide insights on how 'sense of place' may emerge from the analysis of user generated images and how this knowledge may feed into research on place-based identity and memory construction and support decision making around social/cultural provisions in specific locales; on the other hand, this pilot project is expected to experiment with image analysis that goes beyond the technical characteristics of images and their metadata and engages with their content, contributing, therefore, to the question of how data analysis and, more broadly, machines may construct meaning.



# ePrescribing Risk and Safety Evaluation Tool

**Stephanie Klein**

Project Manager

**Medical Sciences**

Faculty

## Overview

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.



# EpiChange

**Peter Taylor**

Principal Investigator

**Science, Agriculture and  
Engineering**

Faculty

## Overview

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.



# FinTrust: Trust Engineering for the Financial Industry

**Aad van Moorsel**

Principal Investigator

**Science, Agriculture and  
Engineering**

Faculty

## Overview

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.





# FRONTCAT: UI for the CityCAT Flood Model

**Chris Kilsby**

Principal Investigator

**Science, Agriculture and  
Engineering**

Faculty

## Overview

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.



# Geosciences CDT Technician

**Stuart Barr**

Principal Investigator

**Science, Agriculture and  
Engineering**

Faculty

## Overview

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.

## Links

<https://research.ncl.ac.uk/geospatial-systems/programme/>



# Hate Speech: Measures & Counter- Measures

**Helen Margetts**  
Principal Investigator

**Science, Agriculture and  
Engineering**  
Faculty

## Overview

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.



# Human Cell Atlas

**Muzz Haniffa**

Principal Investigator

**Medical Sciences**

Faculty

## Overview

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.

## Links

<https://haniffalab.com/>



# Learning Machines

**Alan Wilson**

Principal Investigator

**Science, Agriculture and  
Engineering**

Faculty

## Overview

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.



## Overview

The way we read is in flux. We are in the midst of the 'digital revolution' which challenges the way books are available to their readers and how they are produced, sold, and read. Yet the popularity of printed books and handwritten crafts continues to grow. This AHRC Early Career Leadership Fellow project, Manuscripts after Print c.1450-1550: Producing and Reading Books during Technological Change, took advantage of the current, crucial moment in the history of the book to examine the way in which books were produced and read in the past, the way they are produced and read now, and the way they might be produced and read in the future.

## Links

<https://research.ncl.ac.uk/mssafterprint/>

<https://hands-on-reading.atnu.ncl.ac.uk>

# Manuscripts After Print

**Aditi Nafde**

Principal Investigator

**Humanities & Social  
Sciences**

Faculty

# MERLON

**David Greenwood**

Principal Investigator

**Science, Agriculture and  
Engineering**

Faculty

## Overview

MERLON introduces an integrated modular local energy management framework for the holistic operational optimisation of local energy systems in presence of high shares of volatile distributed renewable energy sources.


As part of a 15-partner consortium spread across 6 countries, researchers at Newcastle University used machine learning and linear programming techniques to forecast and schedule renewable energy generation, battery storage requirements and other flexible energy resources. The outputs from these algorithms are communicated to a central 'orchestrator' via a REST API to manage the local energy systems.

## Links

<https://www.ncl.ac.uk/engineering/research/eee/projects/integratedmodularenergysystemsmerlon>

<https://www.merlon-project.eu/>





# Multi-scale Infrastructure Systems Analytics (MISTRAL)

**Stuart Barr**

Principal Investigator

**Science, Agriculture and  
Engineering**  
Faculty

## Overview

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.

## Links

<https://gtr.ukri.org/projects?ref=EP%2FN017064%2F1>



# NU-LAB: Newcastle University LAB Digital Edition Publishing Cooperative

**James Cummings**

Principal Investigator

**Humanities & Social  
Sciences**

Faculty

## Overview

NU-LAB will create 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. The NU-LAB Project will improve or contribute to containerization of LAB|Pro using docker or similar technologies. All developments, testing, and case studies will be documented in full reports. These reports will improve the generalisability of LAB|Pro and the adoption of it for the creation and publication of scholarly digital editions at other institutions outside the initial scope of American liberal arts colleges.

## Links

<https://research.ncl.ac.uk/atnu/projects/nu-lab/>



# Organ Retrieval Decision Support

**Colin Wilson**

Principal Investigator

**Medical Sciences**

Faculty

## Overview

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.



# Payday Loan Advisor



**Ronnie Das**

Principal Investigator

**Humanities & Social  
Sciences**

Faculty

## Overview

High-cost short-term credit, often referred to as 'payday loans' can provide people with access to large amounts of money after only few clicks and little by way of affordability checks. Informed by a comprehensive journey map of decision making of those considering a payday loan the website will offer personalised advice.



# PlaSMA

**Charles Morisset**  
Principal Investigator

**Science, Agriculture and  
Engineering**  
Faculty

## Overview

We are working with researchers in computing to design, deploy and test a data Platform for Space Monitoring and Analysis (PlaSMA), which will enable space managers to monitor and analyse space policy enforcement and to respond to real-time bespoke alerts. PlaSMA will use privacy preserving data collection and analysis techniques, built on top of existing smart building infrastructure. PlaSMA will also be equipped with a privacy dashboard, co-designed with space users, enabling them to understand how space is used and monitored, and to report on issues and concerns. Finally, PlaSMA will investigate and integrate the legal aspects of privacy-preserving space monitoring digital surveillance, which is still a relatively blank field, to inform space managers.



A detailed anatomical model of a human heart and kidney. The heart is shown in cross-section, revealing its internal chambers and valves. The kidney is positioned below the heart, showing its characteristic bean shape and internal structure. The model is set against a blue background with a faint, stylized pattern of a kidney.

# Quality in Organ Donation

**James Shaw**

Principal Investigator

**Medical Sciences**

Faculty

## Overview

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.



# Rest & Be Thankful Landslide Detection

**Stuart Dunning**

Principal Investigator

**Science, Agriculture and  
Engineering**

Faculty

## Overview

Researchers in Geography are developing methods of landslide detection and mitigation. They have an ideal natural landslide laboratory at the A83 at Rest and Be Thankful in Scotland, a key trunk road where more than £8 million has been spent on landslide mitigation and events occur with regularity, multiple times per year. Previous grants have funded initial sensors, and, a slope-wide resilient 4G / satellite Wi-Fi system to get data to the cloud for live processing. The crux of this project is to be able to efficiently and rapidly bring these data streams that are processed in many different languages and software together, and allow them to be used in combination where they are far more powerful.



# Research Impact Fund: HaSS (RIF)

**Paul Watson**  
Principal Investigator

**Humanities & Social  
Sciences  
Faculty**

## Overview

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.



# Murray Dick

## Arts and Cultures



# Michael Richardson

## Geography, Politics and Sociology

Website application that presents a different way of understanding the evolution of investigative journalism in the UK from conventional, narrative-based histories.

<https://investigativejournalism.ncldata.dev/>

Collaboration involving North East Young Dads and Lads, and Seven Stories, The National Centre for Children's Book. The application includes walking route with Augmented Reality experiences.

<https://digital-storytelling.ncldata.dev/>





## RIF Call 3: May 2020

### **Barbara Gribling**

English Literature, Language  
and Linguistics

### **Fiona Menger, Julie Morris, Matt Forshaw**

Education, Communication and  
Language Sciences &  
Computing

### **Heritage Explorers**

Digital recreation of a popular British history  
and heritage-themed educational board  
game from the nineteenth century. The site  
offers traditional race-based gameplay  
around the board as well as mini-games that  
bring the board to life.

### **Digitised Assessment for Aphasia of Written Naming (DAAWN),**

DAAWN is an assessment of single word  
typed naming designed to be used by Speech  
and Language Therapists with people with  
aphasia.

<https://daawn.ncldata.dev/>



## RIF Call 4: August 2020

**Ronnie Das**

Business School

**Nick Riches, Lauren  
Ackerman, Carol Moxam,  
Cristina McKean, and  
Christine Jack**

Education, Communication and  
Language Sciences & English  
Literature, Language and  
Linguistics

### **Finline**

Website to help small and medium enterprises effected by COVID-19 assess and improve their resilience. Awarded best project in 'Challenge 2: An economy that works for people' at the EU Datathon 2020.

<https://finline.ncldata.dev/>

### **Development of a Teletherapy Platform for Research on Teaching and Learning**

A platform to facilitate therapists to deliver remote interventions to children with Speech, Language and Communication Needs during the COVID-19 pandemic and beyond.



# RIF Call 5: November 2020

**Stuart Dunning**

Geography, Politics and  
Sociology

**Gemma McKinnie**

Modern Languages

## **IoT Sensors in real-time landslide risk reduction**

A cloud based solution to integrate multiple data streams from the natural landslide laboratory at the A83 Rest and Be Thankful in Scotland, to display critical data and trigger decisions.

## **Screening Violence: experiences of conflict around the globe**

Explore video footage from around the world in which participants engage with each other on issues from justice and reconciliation to memory and loyalty.





# Rome Transformed

**Ian Haynes**

Principal Investigator

**Humanities & Social  
Sciences**

Faculty

LATERAN PROJECT

## Overview

Rome Transformed (ROMETRANS) aims to advance our understanding of Rome and its place in cultural change across the Mediterranean World by mapping political, military, and religious changes to the eastern Caelian from the first to eighth centuries CE. The programme offers multiple gains for archaeologists, historians, topographers, and geographers by documenting both the mundane and monumental elements of the city fabric in chronological, geographical, and ideological relationship to one another.

## Links

<https://research.ncl.ac.uk/rometrans/>

# Senescent Cell Counting

**Thomas von Zglinicki**

Principal Investigator


**Medical Sciences**

Faculty

## Overview

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.





## Improving outcomes for women diagnosed with early breast cancer through adherence to adjuvant endocrine therapy (SWEET)

**Linda Sharp**

Principal Investigator

**Medical Sciences**

Faculty

### Overview

There is strong evidence that five or more years of adjuvant endocrine therapy (AET) substantially reduces risks of recurrence and mortality in some forms of early breast cancer. However, many women struggle to adhere to this therapy for such a long period.

SWEET will develop and test a support package to help women adhere to their therapy, centred around an app/website which will: provide information about AET; allow users to monitor their own adherence; and support them to self-manage any problems, symptoms and side-effects that they might experience.



# Smart Cities & Inequality

**Rachel Franklin**

Principal Investigator

**Humanities & Social  
Sciences**

Faculty

## Overview

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.



# Streaming Data Modelling

**Darren Wilkinson**

Principal Investigator

**Science, Agriculture and  
Engineering**  
Faculty

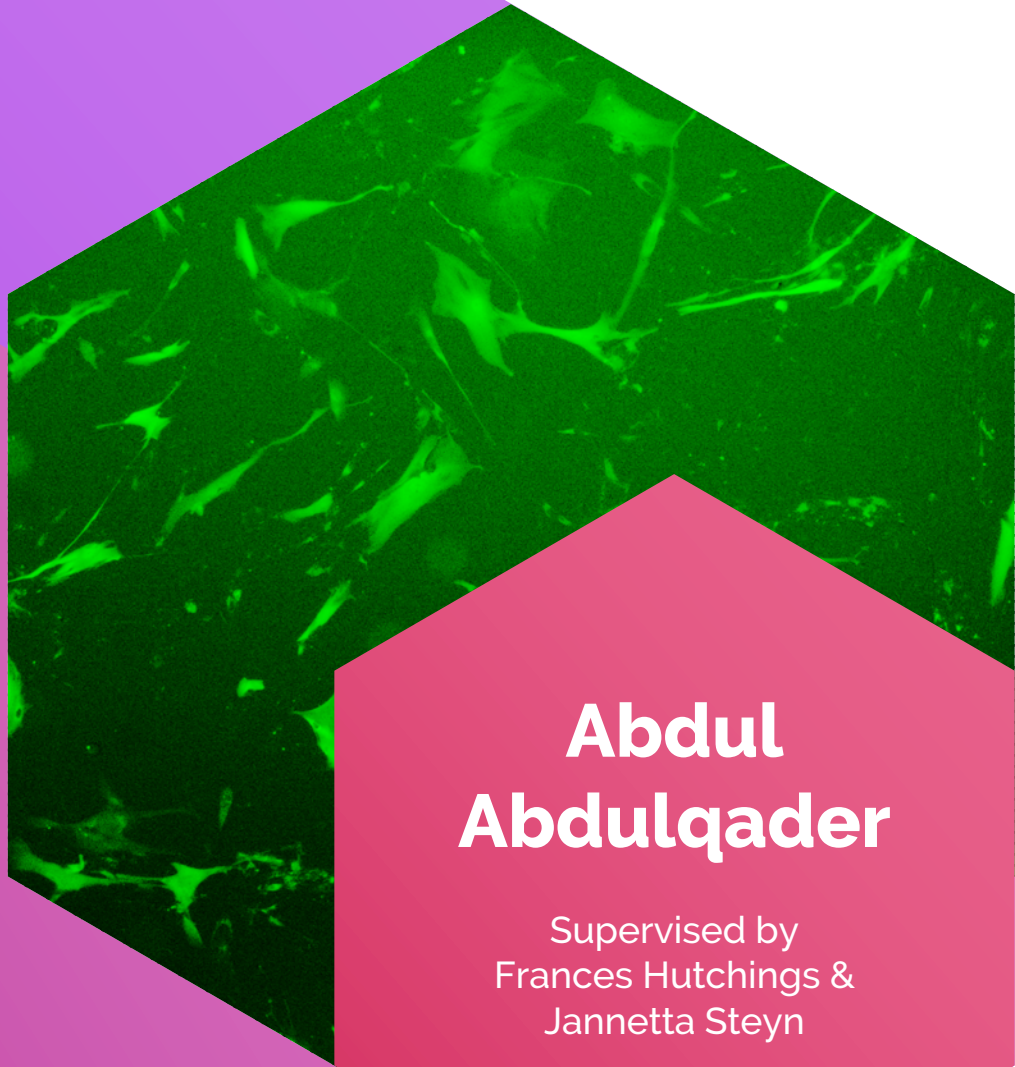
## Overview

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.

# Students

Project contributions by  
our MSc Students





**Abdul  
Abdulqader**

Supervised by  
Frances Hutchings &  
Jannetta Steyn

## Machine Learning for Senescent Cell Counting

It can be difficult to accurately detect the number of cells in a microscope image, particularly for cells with a large variance in shape and size. We were approached by researchers from the Newcastle University Biosciences Institute who had been facing this issue in the context of evaluating senolytic (anti-ageing) drugs. This student project aimed to improve upon initial work by the RSE team to create a machine learning solution to this problem. The project trialled different machine learning approaches and scanned model parameters to create a functional cell counting solution that can be deployed in the ageing research labs, and potentially beyond.

# Sam Colenutt

In a collaboration with the Newcastle University Biosciences Institute, this project involved working to improve a machine learning model for counting cells in microscope images. The project focused on the problem of dealing with small datasets as annotation of cells in large microscope images is very time consuming. The student explored different ways to artificially extend a dataset of annotated images, using transforms and image splitting techniques.

## Sam Colenutt

Supervised by  
Mark Turner &  
John Brennan (NICD)







**Cristina  
Deligiorgi**

Supervised by  
Frances Hutchings &  
Jannetta Steyn

## Containerisation of an Automated Cell Counting Model

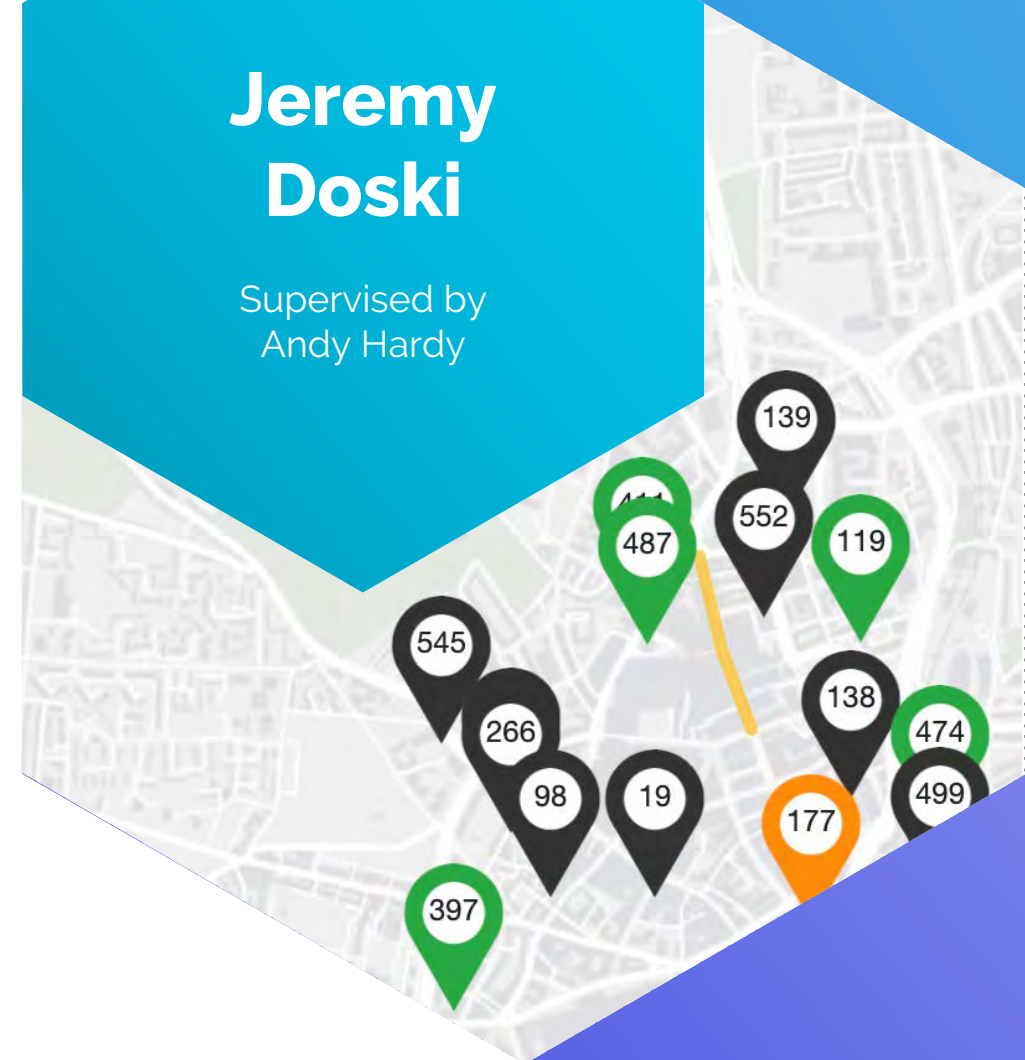
It can be difficult to accurately detect the number of cells in a microscope image, particularly for cells with a large variance in shape and size. We have a machine learning algorithm, implemented in Python, trained on microscopy data which can output an estimated count. However, the current results are still not as accurate as we would like. This project will involve improving the current code in order to increase the counting accuracy. The student will learn about machine learning through working on this real-world application, which, if successful, will be used in the biology laboratories at Newcastle University.

# Jeremy Doski

A project to enable users to find their closest free parking space. The application will make use of the urban observatory data stream to visualise parking availability on a map in real-time. The application should work on all the standard devices from desktop to mobile but an emphasis should be made for mobile-first. If time allows enhancement could be to enable the application to use audio to notify the user of their nearest parking space and integrate with direction APIs to plot a route.

## Jeremy Doski

Supervised by  
Andy Hardy







## Scott Hamilton

Supervised by  
Kate Court

# The Books of The Bees

A table-book has each part of a piece of music printed facing a different direction so that performers can sit surrounding the book. This project involved creating a web application that enables people to sing together with each device displaying one part of the music. The student employed web sockets to address the challenge of co-ordinating music across devices. After a user creates an ensemble they can invite others to join and select their parts. When the host starts play, individual bars of music scroll at the same speed across devices so users can sing in unison. When accessed through a phone or tablet the long-term aim is that people can arrange their devices on a table in front of them to echo the arrangement of music in a table-book. This project was developed into a pilot application as part of the ATNU project.

# Antarctic Coastline Mapping

The Antarctic is a rapidly changing environment. Many of the key geospatial datasets essential to support research and logistics in Antarctica (for example the coastline) need updating frequently. This has thus far been done manually, at considerable resource cost. The MSc project offered explored the possibility that much of this digitisation work could be undertaken by automatic algorithms, with expert feedback helping digitisation software to learn the process and refine the process. Anastasija Jadrevska investigated in depth how readily-available satellite imagery could be used as input to a software pipeline to automatically identify ephemeral and sometimes chaotic ice coastline fronts. Her work evaluated the usefulness of several edge-detection algorithms, and will provide a valuable basis for further research projects in future.



Anastasija  
Jadrevska

Supervised by  
David Herbert





**Monika  
Kucinska**

Supervised by  
Frances Hutchings

## Extending Annotated Datasets of Senescent Cells

In a collaboration with the Newcastle University Biosciences Institute, this project involved working to improve a machine learning model for counting cells in microscope images. The project focused on the problem of dealing with small datasets as annotation of cells in large microscope images is very time consuming. The student explored different ways to artificially extend a dataset of annotated images, using transforms and image splitting techniques.

# COVID-19

Contributions to the  
University's response to  
the COVID-19 pandemic





## CPAT

The team worked alongside colleagues from Computing to create a tool to help clinical staff at Newcastle's Royal Victoria Infirmary prioritise COVID-19 patients.



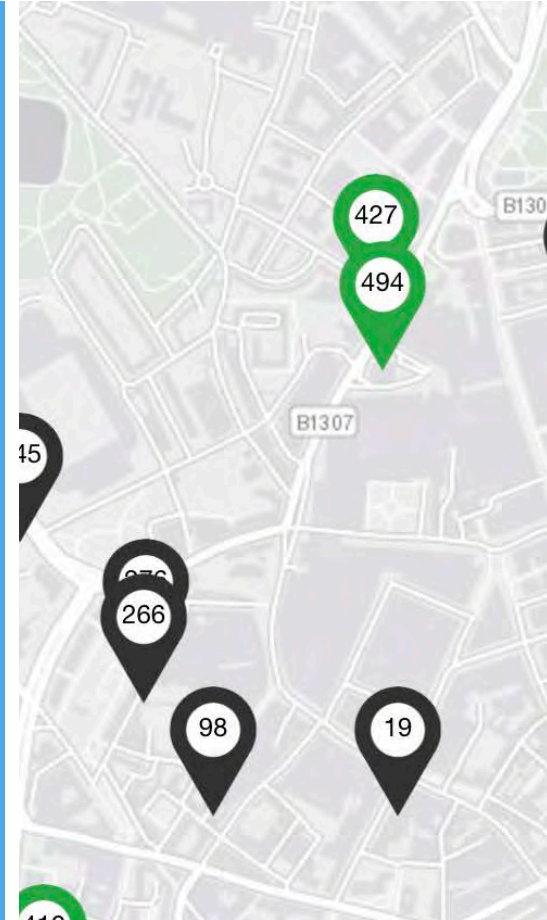
## IDEA-FAST

We also picked up work in the Medical School to support teams with reduced staff. For IDEA-FAST we helped with data entry for automated database design.



## Mutual Support App

Many University staff volunteered in their local community in the first lockdown. We helped an academic collate public address data for organising volunteers by postcode.



## NE1 City Dashboard

We created a proof-of-concept app for Newcastle City Council called [howbusyistoon.com](https://howbusyistoon.com). The site showed the public how busy the shopping streets and car parks were based on data from Urban Observatory.



# Testimonials

Experiences of working  
with the team from  
academics



**David Greenwood**

NU Academic Track Fellow

Faculty of Science, Agriculture &  
Engineering

# David Greenwood

As part of the **MERLON Horizon 2020 project**, Newcastle University researchers were tasked with producing forecasting and scheduling modules for island local energy systems. The project aims to make better use of renewable energy systems and improve reliability to customers by coordinating local resources and enabling the local network to operate as an energy island. This required a scheduling approach which could keep the local energy resources in balance and forecasting methods to predict the short-term future behaviour of the system. While the researcher had the skills and knowledge to develop these algorithms, they were not equipped to produce robust implementations or to integrate the algorithms with software modules developed by other partners.

The RSE team worked effectively with us to fill these gaps, from initial conversations with Mark Turner, through to Jim McGrath reimplementing the researcher's code and creating the necessary data interfaces. The project would have been extremely challenging for us without RSE support. Jim was very easy to work with, gave clear communication, explaining what he was going to do as well as why it was the best approach and how it would work. He has been available to answer questions and provide technical input since his formal time on our project ended. Next year, we look forward to working with Jim again as we deliver an updated version of the modules informed by trial results.



# Muzlifah Haniffa

The Human Cell Atlas (HCA) is a global initiative that aims to create a reference atlas of the human body by mapping every cell. Having a reference atlas of the healthy human body will allow us to understand what has changed in disease and how we can reverse or ameliorate disease processes. The Haniffa Lab contributes to the Human Developmental and the Skin Cell Atlas using single cell and spatial genomics approaches to map the cells during human development and in the skin.

A key component of HCA is provision of the atlas data in repositories and access through webportals for any user to browse and query the data. We realised the increased software engineering expertise required to maintain and enhance a prototype webportal that had been established by a post-doctoral fellow in the lab. We formed a partnership with the Newcastle RSE team and recruited two RSEs, Dave Horsfall and Jim McGrath, for our project. Despite the pandemic constraints and lack of face-to-face contact, we have worked together successfully to integrate Dave and Jim across the RSE team and the HaniffaLab. The collaborative exchange and software engineering input has resulted in significantly enhanced features to our existing web portals as well as new data queries e.g. disease causing genes within the cell atlas data visualisation options.

Building on from the initial success, Dave and Jim are currently preparing an entirely new web portal infrastructure that is more robust, scalable and with enhanced features including basic analysis of the data in real time. We aim to make the web portal intuitive, fun and informative for a broad range of scientific, clinical and lay audience.



**Muzlifah Haniffa**

Professor of Dermatology  
and Immunology

Faculty of  
Medical Sciences



**Ian Haynes**

Professor of Archaeology

Faculty of Humanities  
& Social Sciences

# Ian Haynes

Aided and abetted by advances in architectural modelling software, it is now possible to generate photo-realistic images of ancient structures as they are believed to have originally appeared. While some representations are based on exhaustive research and can be robustly justified to the smallest detail, others owe more to the imagination of their digital creator than to scientific evidence. The ERC Advanced Grant **ROME TRANSFORMED** (PI Prof Ian Haynes) confronts this problem by seeking to make the reasoning behind its digital models not merely transparent, but also integral to scholarly dialogue into the form of function of ancient structures.

Drawing on the work of Dr Marc Grellert of Darmstadt Technical University, creator of SCEIDOC, a system designed to link Scientific Documentation with visualisations, Dr Mike Simpson and the RSE team have worked with both Marc and ROME TRANSFORMED team to generate a new user-friendly system that will perform all these functions. Mike has worked closely with members of the team to modify and augment the SCIEDOC software to create a solution that will greatly benefit the ROME TRANSFORMED project, as well as contributing additional features to SCIEDOC itself.



A silhouette of a rock climber is visible on the left side of the slide, rappelling down a rock face. The climber is wearing a helmet, a harness, and a rope. The background is a solid dark red color with a large, faint, stylized letter 'C' in the center.

# Challenges

Looking ahead to some of  
the challenges the team  
will face in 2021

# Challenges

Looking ahead to 2021 and beyond, the team faces a number of challenges. Here, we focus on the three most pressing issues.

## Facility Transition

Transitioning from a model of costing staff onto grants to a service-oriented facility that charges a day rate.

01

## Diversification

Positioning ourselves in a way that we can take on work beyond the standard research projects.

02

## COVID-19

Adjusting to a post-pandemic world in both academia and changes to working practices.

03



# Facility Transition

Transitioning to a new funding model

Our biggest challenge last year was running the RSE model within the confines of existing University bureaucracy. Finance and project support services are setup in a way that helps research groups working on a small number of long running grants. Trying to frequently move staff between grants on fractional splits becomes difficult in such setups. The answer to this is to run the team as a research facility that recovers costs by charging for time as a service, rather than individual RSEs. In the long term, this will significantly reduce the time need to administer the team but things will continue to be complex in the short term.

We have existing projects either in active development or submitted to funders and those costings must be honoured. This means the direct costs model will run for a few more years and be phased out as each team member moves to projects costed using the facility day-rate model. The next 8-12 months will be the most chaotic but there will also be a long tail – our longest current grant has another 7 years to run.

The facility will have a flat rate that builds in some utilisation buffer, allowing gaps between projects or for individuals to be slightly under utilised over a year. This increases job security and ensures stability. Over time, as RSEs become 100% funded by the facility, employment will become tied to it's existence and not individual grants. Financial reporting will become much easier, improving on the complex and laborious task of pulling together grant information from different finance teams. Having a full clear financial outlook means strategic decisions can be made with greater confidence, allowing us to make smarter decisions in areas such as recruitment, training and purchasing.

Whilst the new model will solve some problems, it comes with some challenges. The comparative day rate will rise in line with staff members coming out of the University overheads calculation, making the service slightly more expensive will need to be backed up with an increase the breadth of services the team offers, using our new found flexibility to do more to support research and justify the decision to move to this model.

# Diversification

Widening our offer to researchers

Once the move to our new funding model is complete, focus will shift to what other services we can provide that will help researchers, others in RSE-like roles and the University more generally. Running training courses in the fundamentals of software engineering is a first step towards these goals. The whole team are now trained Software Carpentry instructors but we can broaden the range of subjects we cover. Demand for places is already high, especially for the Introduction to Python and Version Control courses. Balancing the main activities of the team with meeting demand for training will be challenging.

Some of the requests for support we get are from people who are already competent programmers, so introductory training is not suitable route to support them. One possible service could be an application or architecture design session. These would be whiteboard or discussion sessions with the applicant to think through various challenges with their code and how those might be solved. The key different here is that the applicant would then go away and build that solution themselves.

The University has made a concerted push over the last few years to advocate for open research as standard. This is led by the Data Service based in the library, which maintains an open data repository for data repository for the storage of research datasets that can be used by others. The RSE Team could be doing much more to integrate our projects with this service, both in terms of adding data into the repository and utilising other datasets held within it.

Finally, there are now a wide range of RSEs or RSE-like roles around the University. These are either embedded roles within specialist research groups or other technical roles with significant software aspects. The Team should be doing more to support the professional development and technical support for these individuals. This could be anything from a place to bounce ideas around, if an individual is the lone software engineer in their group, or more structured support roles with regular events and opportunities to meet other software engineers.



# COVID-19

Adjusting to impacts of the pandemic

Like everyone else, we have had to adjust to a 100% remote working pattern. Previously, team members could work from home whenever they wanted, but most took advantage of being in the office at least a couple of days a week and some were 100% office based. Adapting to remote working has not been easy and is likely to continue in some form well into 2021. Collectively, we need to ensure we are providing enough support to colleagues who prefer office working as well as being understanding of the other impacts of remote working such as childcare/home schooling during lockdowns, caring for family members and impacts from self-isolation.

Given the nature of the pooled resource that the team offers, it is vital to keep up efforts to bond the team together. This is much harder when working remotely, but is not impossible. Events that replicate ad-hoc office based moments, such as lunch and coffee breaks, have become contrived but still valuable moments to connect with colleagues and check-in with home life. There is multitude of tools that have emerged that make it easier

to do these things remotely, but none of them fully capture the in-person experience and some have other downsides. The idea of 'Zoom fatigue' is much discussed, but trying to avoid placing unnecessary burden on each other with expectations to make every interaction a video call is a tough balance to strike.

During the first lockdown in March 2020 there was an increase in the number of requests for *pro bono* work for efforts related to the pandemic. These were either in direct support of medical or healthcare efforts or to support research and teaching that had to be moved online. As the year has gone on the requests have become less frequent but do still happen, responding to them in our current funding model is difficult when the instinct is to be as helpful as possible.

# Opportunities

Areas we can contribute more in 2021



# Opportunities

As well as challenges there are also many opportunities for us to do more to support the University and wider RSE work.



01

## **Raising Awareness**

We can do more to support colleagues in similar roles and help with career development and training.

02

## **Newcastle Data**

Leverage the combined skills and experience of the wider group and support new initiatives.

03

## **Student Projects**

Learn the lessons from our MSc projects and improve our student experience.

# Awareness Raising

Supporting RSE roles

Outside of the central RSE team, similar roles are appearing around the University. In some cases these are Research Software Engineers, but there are also other technical roles that have significant software engineering components. The team should position itself to support those who are working in isolated roles and form relationships with other groups and individuals that may be mutually beneficial. For those working in more isolated roles, restarting our Code Community group will help connect people with others doing similar roles. For larger teams, seeking opportunities to work on larger research projects together would create an environment for closer collaboration in future.

For individual software engineers, regardless of if they're working alone or part of another team, an informal mentoring scheme might be useful. The University already has a tried and tested mentoring scheme, however the scheme is rightly focused on professional and career development and this mentoring scheme would focus on technical skills. By keeping it informal, individuals would be matched with someone from the RSE team who has

knowledge of the technical skill they're trying to learn. It would be a place to turn for advice and guidance on methods to improve their programming ability.

One-on-one mentoring will not scale well, so for larger needs a formal training programme and signposting will be needed. Our training materials are currently pitched at a beginner level, but could be expanded to cover more advanced topics. We should also be making much more of our connections outside the University when it comes to advanced training, either by advertising existing courses or seeking to bring in specialists to run events. We're well positioned within the Alan Turing Institute network and N8 network to do this.

To bring all of this together we could run or sponsor an internal or regional RSE mini conference focused around practical sessions for training and mentoring as well as showcasing the software engineering work being done around the University.



# Newcastle Data

Working with our wider group

The RSE team is based in a wider group covering research, teaching and engagement activities that broadly cover applied computer science and data science. Communication between these elements is good, but we could be doing a lot more to work more collaboratively with each element.

Research is covered by the Scalable research group in the School of Computing and the Data NUCoRE, a cross-disciplinary group of academics looking at applied data science. We already work with a number of these academics on individual standard RSE projects, but could look to play a more strategic role within the research space. The Data NUCoRE particularly is a route to wider networks of academics that we could collaborate with.

Newcastle Data works with the School of Computing to operate a Centre for Doctoral Training in Cloud & Big Data and an MSc in Data Science. These activities broadly cover its teaching activities. Some of our MSc projects have been completed by students from the MSc, but we are yet to work on anything specific in partnership with the CDT programme. There

are opportunities with the MSc to reserve places or run private sessions to deliver software carpentry skills training as well as continue to offer projects to those students. With the CDT, collaboration is less obvious due to the individual nature of working on a PhD. There may be opportunities to work alongside the student and their supervisor on side projects that often emerge from PhD research.

Newcastle Data's outreach activities are spearheaded by the National Innovation Centre for Data (NICD). NICD works with national businesses to upskill their workforce by jointly working on data science projects that are brought to NICD by the business. Although the missions are different, there is a great deal of overlapping skills between the NICD technical team and the RSE team. There is scope to work together on large scale research projects that involve industrial partners, and we should continue to seek out such opportunities.

# Student Projects

Improving student experience

Last year, we offered a small number of MSc Computing Projects that proved to be very popular with students. The projects were made available to all MSc Computing students regardless of what type of course they were on. Due to the way we're funded in comparison to academic staff in the school, we made the decision to cap numbers to one student per project and RSEs offering no more than two projects. We met with large numbers of students interested in the projects and then selected those with the best skills fit and those most enthusiastic about the research topic. When meeting students, the popularity of the projects was due to the real-world nature of the problems and the chance to work as part of a team to deliver something at the end.

Pre-COVID we had made space in the office to allow students to come and work alongside the team but by the time they started we were in the first national lockdown. We would like to try this again in 2021 or 2022 as we feel students will get a lot from treating the project like a work placement

whilst still delivering their final year project. In addition to being co-located with the team, they would also be with each other, creating a cohort experience that hopefully they use to start helping each other.

The feedback we received from students graduating in 2020 was excellent, many have asked to stay in touch and be kept informed of job openings in the future. We have a real opportunity to do more in this area, and figuring out ways to scale our offering to students without jeopardising the experience will be part of our planning in the next few years. We need to be mindful of spreading load around the team, keeping our project ideas high quality but suitable for the level of the students and ensuring we don't overstretch ourselves and lose track of our core purpose. If we can do those things there is no reason we can't offer a better experience to students.