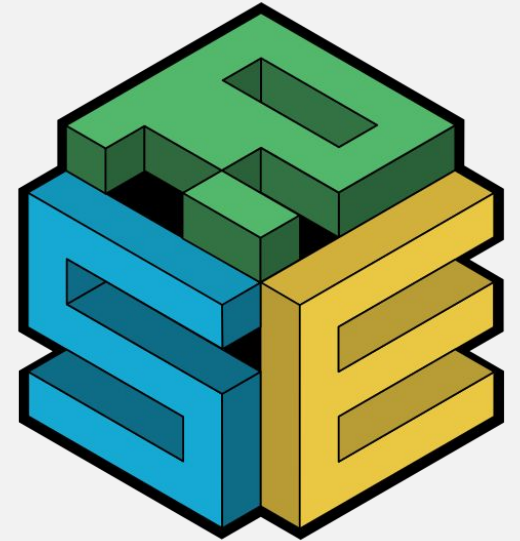




The
University
Of
Sheffield.

Research Software Engineering at Sheffield



Q&A Jamboard: <http://bit.ly/RSE2021QA>

A WORLD
TOP 100
UNIVERSITY



The
University
Of
Sheffield.

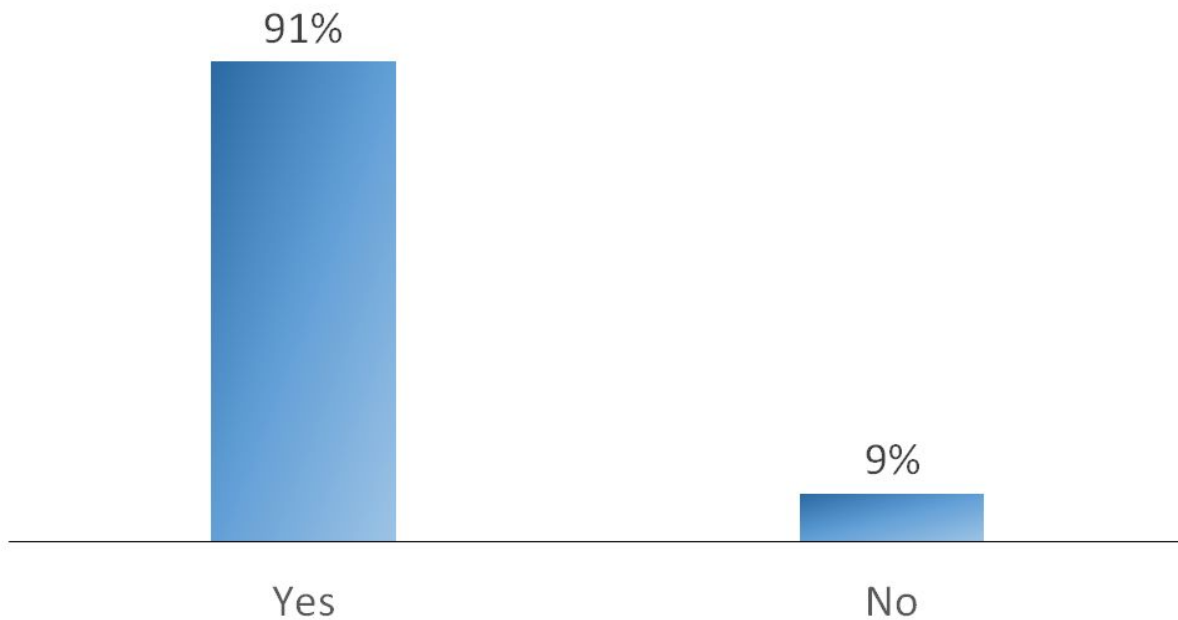
SOFTWARE IS
ESSENTIAL FOR
RESEARCH

A WORLD
TOP 100
UNIVERSITY



The
University
Of
Sheffield.

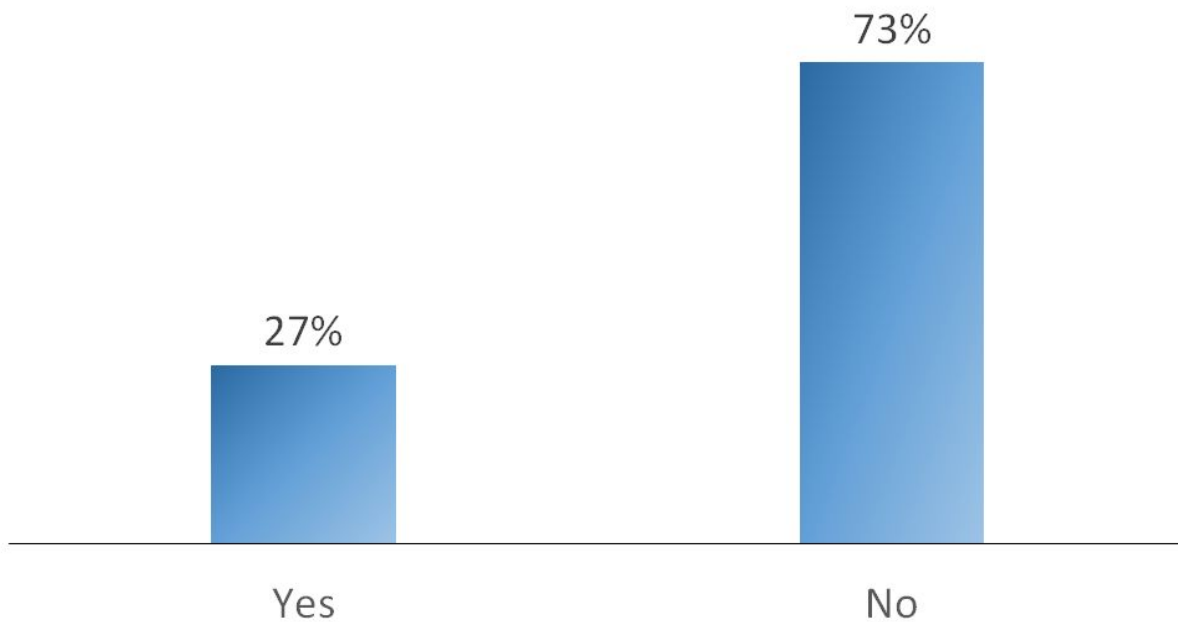
Do you use research software?
(382 responses)





The
University
Of
Sheffield.

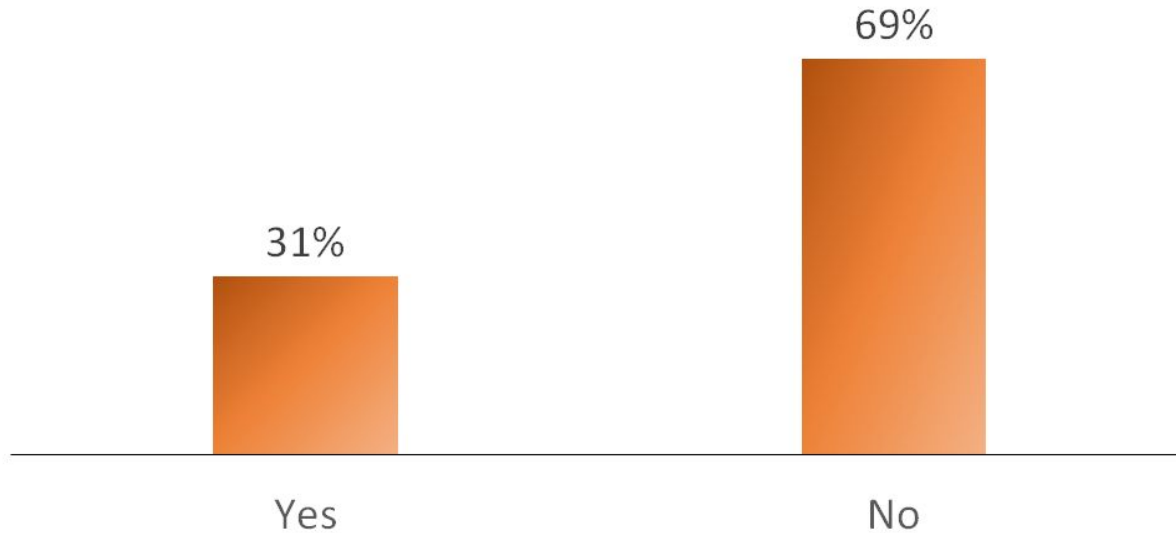
Do you develop research software?
(382 responses)





The
University
Of
Sheffield.

Do you feel you have received sufficient
training to develop reliable software?
(103 responses)





The University Of Sheffield.

A Scientist's Nightmare: Software Problem Leads to Five Retractions

Greg Miller
+ See all authors and affiliations

Science 22 Dec 2006:
Vol. 314, Issue 5807, pp. 1856-1857
DOI: 10.1126/science.314.5807.1856

Article Figures & Data Info & Metrics eLetters PDF

Until recently, Geoffrey Chang's career was on a trajectory most young scientists only dream about. In 1999, at the age of 28, the protein crystallographer landed a faculty position at the prestigious Scripps Research Institute in San Diego, California. The next year, in a ceremony at the White House, Chang received a Presidential Early Career Award for Scientists and Engineers, the country's highest honor for young researchers. His lab generated a stream of high-profile papers detailing the molecular structures of important proteins embedded in cell membranes.

Then the dream turned into a nightmare. In September, Swiss researchers published a paper in *Nature* that cast serious doubt on a protein structure Chang's group had described in a 2001 *Science* paper. When he investigated, Chang was horrified to discover that a homemade data-analysis program had flipped two columns of data, inverting the electron-

Full Text
Help

A WORLD
TOP 100
UNIVERSITY



The
University
Of
Sheffield.

We need research software to be:

- **Correct** tested, testable
- **Understandable** transparency, sustainability
- **Reproducible** research methods & ethics
- (Re) **Usable** (unlocking knowledge)
- **Easy to change** (new knowledge!)

Why is this so hard?



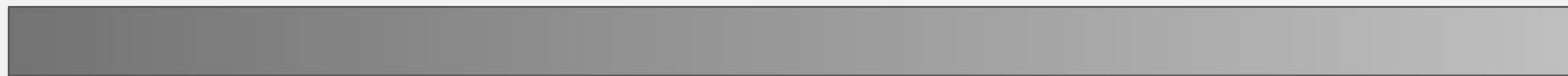
The University Of Sheffield.

Software Engineer

???



Researcher



Research Developer

A WORLD
TOP 100
UNIVERSITY

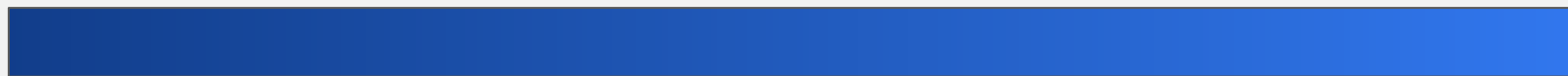


The University Of Sheffield.

The “*Research Software Engineer*”

Software Engineer

Researcher



SOCIETY OF RESEARCH SOFTWARE ENGINEERING

Research Developer

A WORLD
TOP 100
UNIVERSITY



The
University
Of
Sheffield.

What do RSEs do?

- Build **quality** software - as **part of** research projects
- Help researchers:
 - Learn computing skills
 - Establish good practices
 - Navigate computing and data infrastructure

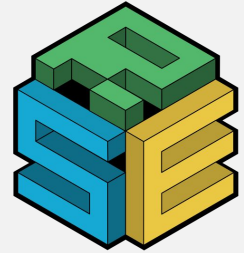


Photo by [Goutam1962](#), under CC-BY



The
University
Of
Sheffield.

Why RSE Sheffield?



For RSEs

- Stable careers
- Peer support
- Recognition & reward
- Professional development

For Research Projects

- Fractional support
- Flexible access to expertise
- Access to niche skills
- Skill sharing between projects

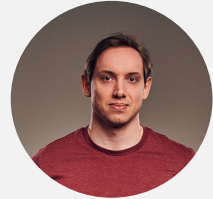
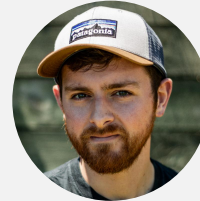
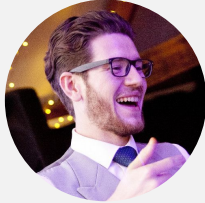
For Researchers

- Help, support and mentoring
- Formal training
- Reproducible process and best practice



The
University
Of
Sheffield.

RSE Sheffield



<https://rse.shef.ac.uk/contact/team/>

A WORLD
TOP 100
UNIVERSITY



The
University
Of
Sheffield.

Some projects we've been involved with

1. **Accelerating software** using Graphical Processing Units (GPUs)
 - Rob Chisholm
2. **Epidemiological modelling software**
 - Bob Turner
3. **Machine Learning and AI**
 - Twin Karamkharm
4. **Impact from software**
 - Will Furnass (in place of David Wilby)

Improving Performance using GPUs

For tumor modelling



English

Search

Home Primage Project Partners Resources Events and News Contact Us

News

[PRIMAGE project M26 Consortium Meeting & Workshop](#)

[New paper in the scope of the PRIMAGE Project](#)

[PRIMAGE Project and CONNECT University](#)

[More news...](#)

PRedictive In-silico Multiscale Analytics to support cancer personalized diaGnosis and prognosis, Empowered by imaging biomarkers

PRIMAGE is one of the largest and more ambitious European research projects in **medical imaging, artificial intelligence and childhood cancer**.

Technical Partners



Instituto de Instrumentación para Imagen Molecular



Universität Konstanz



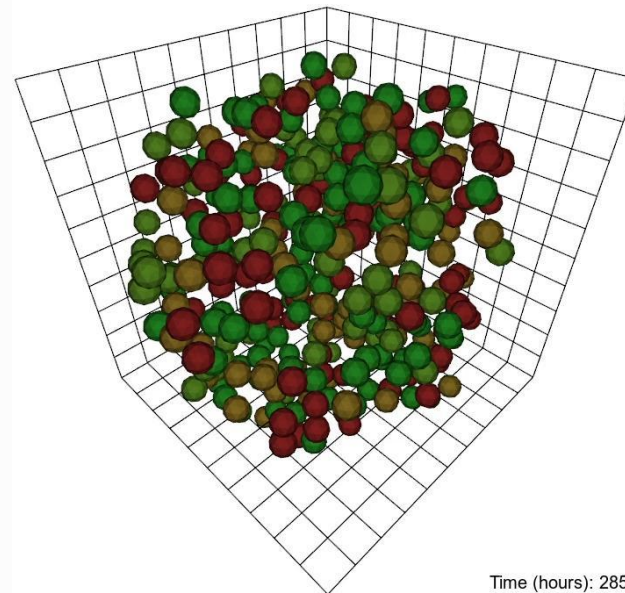
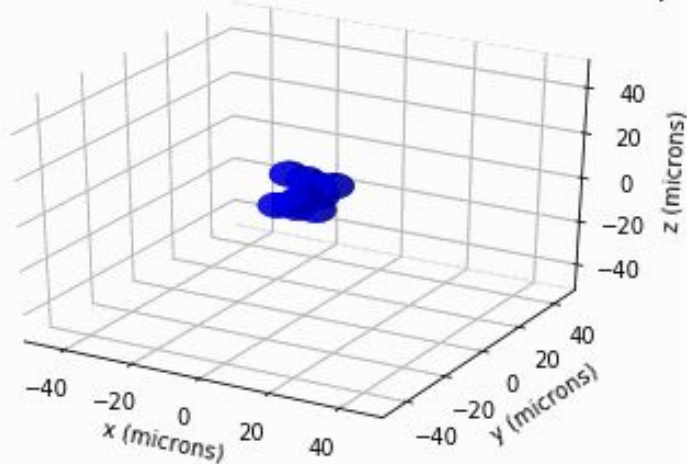
Data Analysis and Visualization





What were the impacts?

Tumour visualisation, after time step 0



Time (hours): 285
Cell count: 382



The University
Of
Sheffield.

#2

THE ROYAL SOCIETY

Venue hire

Home Fellows Events Grants, Schemes & Awards Topics & policy Journals Collections About us

April 2020:
How could we help?

Rapid Assistance in Modelling the Pandemic: RAMP

Fellows of the Royal Society
Coronavirus COVID-19.

See latest activity

The Royal Society has used its
pandemic and guide the UK's
brought modelling expertise from
already working on Coronavirus

An initial call for volunteers led

RAMP is designed to; provide s
used to inform the work of the
apply knowledge from related
information flows.



SCRC docs

Search SCRC docs

Data

Available data
Data Sources and Products
Model Parameters

SCRC data pipeline

Data Registry
Julia
R
Python

Research

SCRC

Who are we?

The Scottish COVID-19 Response Consortium is formed of [dozens of individuals from over 30 academic and commercial organisations](#).

Researchers in these organisations jointly responded to a call by the [Royal Society](#) to develop more epidemiological models of COVID-19 spread - [RAPID ASSISTANCE IN MODELLING THE PANDEMIC: RAMP](#) - in order to develop a more robust and clearer understanding of the impacts of different exit strategies from lockdown. Scientists from several other organisations across the UK and abroad have now joined the consortium to provide additional expertise in specific areas.

A WORLD
TOP 100
UNIVERSITY



The University Of Sheffield.

#2

NEWS WEBSITE OF THE YEAR

The Telegraph Coronavirus News Politics Sport Business Money Opinion Tech Life Style Travel Culture

Gadgets Innovation Big tech Start-ups Politics of tech Gaming

COMMENT

Neil Ferguson's Imperial model could be the most devastating software mistake of all time

The boss of a top software firm asks why the Government failed to get a second opinion from a computer scientist

DAVID RICHARDS AND KONSTANTIN BOUDNIK
16 May 2020 • 1:22pm

nature > news > article

a nature research journal

MENU **nature** Subscribe Search Login

NEWS · 08 JUNE 2020

Critiqued coronavirus simulation gets thumbs up from code-checking efforts

Influential model judged reproducible – although software engineers called its code 'horrible' and 'a buggy mess'.

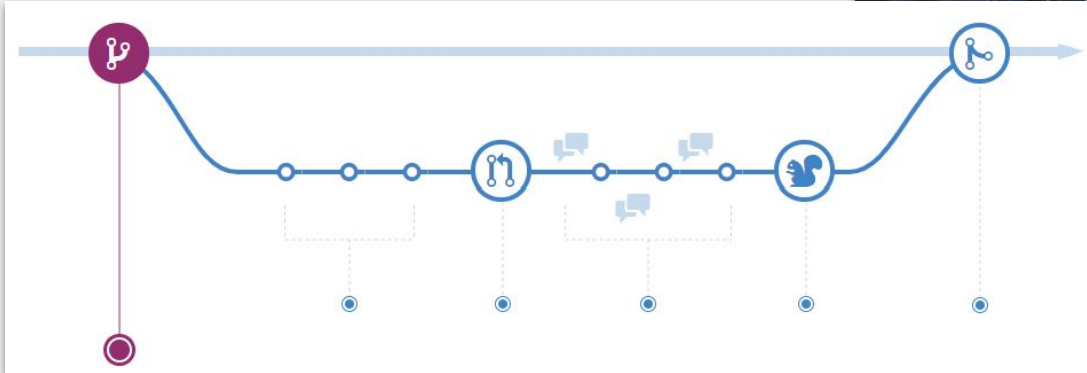
Dalmeet Singh Chawla

A WORLD
TOP 100
UNIVERSITY



The University Of Sheffield.

What did RSE do?



<https://guides.github.com/introduction/flow/>

A WORLD
TOP 100
UNIVERSITY



What were the impacts?

SCRC Software checklist

This checklist has been developed by the Scottish Covid-19 Response Consortium and has been used to assess the software engineering aspects of "model readiness" for our six epidemiological models. This is intended to be part of a broader scheme to evaluate and describe suitability of model results to be used in decision making so this checklist covers only the software implementation and assumes that other documents cover questions about model validation, quality of science, data provenance and quality and policy readiness.

Software Details

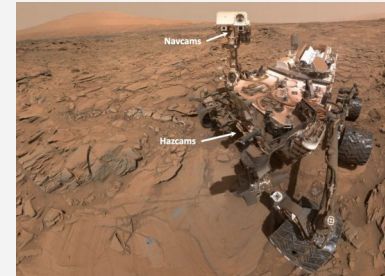
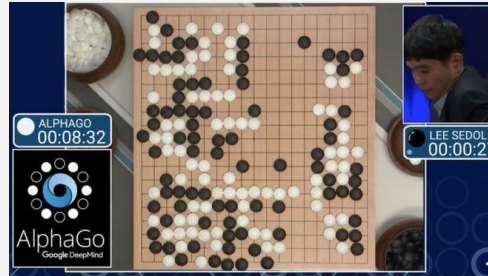
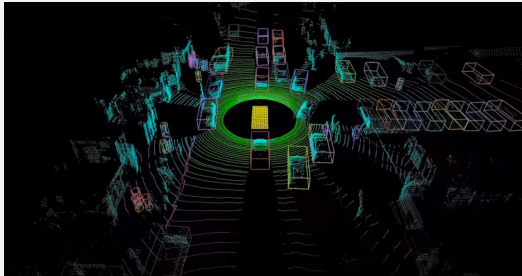
Model / software name

7 pieces of modelling software; rapid reproducible data pipeline; follow on grant



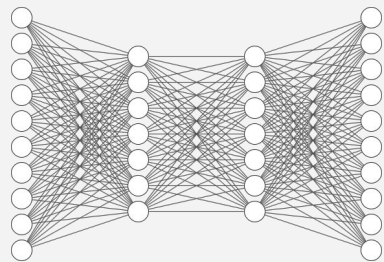
Machine Learning (ML) and Deep Learning (DL)

- Statistical **models** that learns from **data**.
- DL is a sub-category of Machine Learning that uses 'deep' Neural Networks.
- Has huge impacts across most fields in industry and academia.





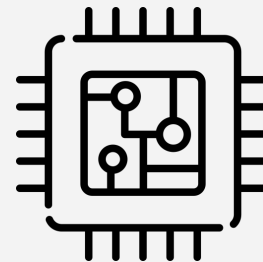
ML & DL: What's involved and how can RSEs help?



**Algorithms and
Workflow**



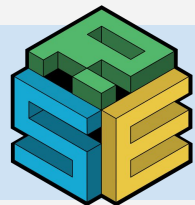
**Data Collection,
Curation and Pipelining**



Computation Resource

Training, consultancy, development

**Hardware consultancy,
HPC support &
on-boarding**





ML & DL: Training

- Introduction to Deep Learning
 - 1-day workshop with Google Colab
 - Practical labs using Tensorflow Keras
 - Available in Python and R

The screenshot shows a Google Colab notebook interface. On the left is a 'Table of contents' sidebar with a search icon and expand/collapse icons. The table of contents includes sections for copyright, license, lab 03 (Image Classification with Convolutional Neural Networks), installing dependencies, exploring data, preprocessing data, and building the model. The 'Build the model' section is expanded, showing exercises 3.1 through 3.6, solutions, and a notice section.

The main notebook area shows a code cell with the following Python code:

```
[9] plt.figure(figsize=(10,10))
i = 0
for (image, label) in test_dataset.take(25):
    image = image.numpy().reshape((28,28))
    plt.subplot(5,5,i+1)
    plt.xticks([])
    plt.yticks([])
    plt.grid(False)
    plt.imshow(image, cmap=plt.cm.binary)
    plt.xlabel(class_names[label])
    i += 1
plt.show()
```

Below the code is a 5x5 grid of image predictions. Each image is shown in a small square with a label underneath. The labels are: Coat, Coat, Ankle boot, Sneaker, Sandal, Trousers, T-shirt/top, Sandal, Sneaker, Coat, T-shirt/top, Bag, Pullover, Dress, Ankle boot, T-shirt/top, Sneaker, Sneaker, Pullover, Pullover, T-shirt/top, Coat, Coat, Coat, Pullover.



The University
Of Sheffield.

ML & DL: Projects

- Rumour Veracity - DL-based service for classifying tweet misinformation and stance
 - Development and deployment of service
 - Twitter rumour veracity classification model implemented in PyTorch
 - Model-retraining infrastructure

Twitter Rumour Verifier

1214812607447736321

Verify!

Results

Tweet

BBC News (World) @BBCWorld

Iran has launched a missile attack on air bases housing US Forces in Iraq

Two air bases were struck in retaliation to the US killing of General Qasem Soleimani

It's unclear whether there have been any casualties

bbc.in/3SzYd4O

7:34 AM · Jan 8, 2020

2K 1.1K

Copy link to Tweet

Your feedback

You can provide your own feedback to the veracity classification here. Drag the slider to specify the veracity and please provide a reason for your classification.

Please drag slider to provide a veracity rating:

False Undecided True

Please provide evidence for the veracity rating:

Submit feedback

Cancel

False Undecided True

Replies

Sort from: Newest first

Show stance: All

Accounts to show: All accounts

Replies to show: 5

Whoops! We couldn't access this Tweet.

Language: es

Richard @richard1985

replying to @BBCWorld and @Albanery

User profile

BBC News (World) (BBCWorld)

User verified

Location: London, UK

Profile Description: News, features and analysis from the World's newsroom. Breaking news, follow @BBCBreaking, UK news, @BBCNews, Latest sports news @BBCSport

Account Created: 01/02/2007 07:44:29

Account Age (days): 5144 days

Followers: 36467150

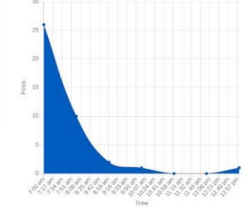
Friends: 71

Number of tweets: 303925

Average tweet per day: 59.083

Request additional user metadata

Replies Histogram



External Links

- In Tweet**
- <http://bbc.in/3SzYd4O>
 - https://video.twimg.com/amplify_video/12148123360972/tag=13
 - https://video.twimg.com/amplify_video/12148123360972/tag=13
 - https://video.twimg.com/amplify_video/12148123360972/tag=13
 - https://video.twimg.com/amplify_video/12148123360972/tag=13

- In Replies**
- <https://pbs.twimg.com/media/ENxrtfQWwAAUuq.jpg>
 - https://video.twimg.com/tweet_video/ENxvQ4XXAACM



The
University
Of
Sheffield.

#3

ML & DL: HPC Support

- Computing resource consultation
- HPC
 - User onboarding
 - Software support
 - Technical consultation
 - Associated systems:
 - Sheffield
 - ShARC
 - Bessemer
 - Tier-2
 - JADE
 - JADE II
 - Bede





Research Software Impact Enhancement



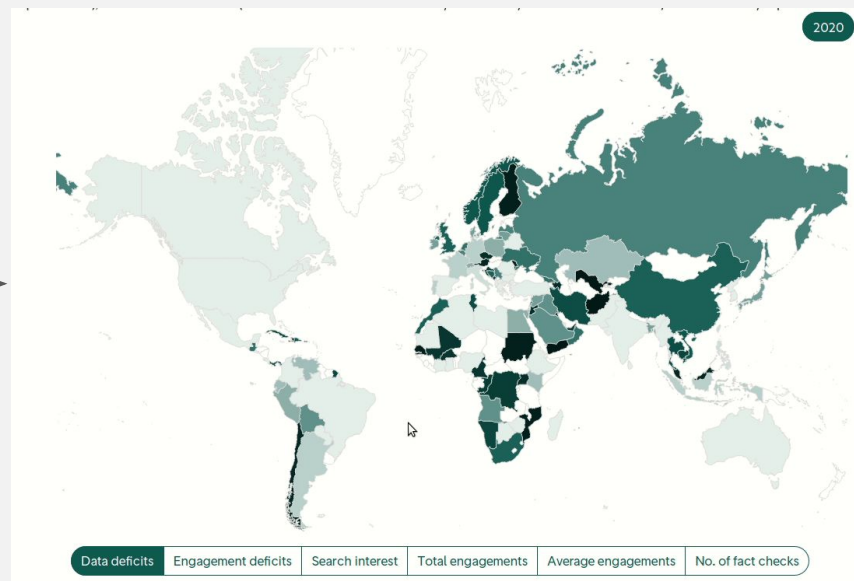
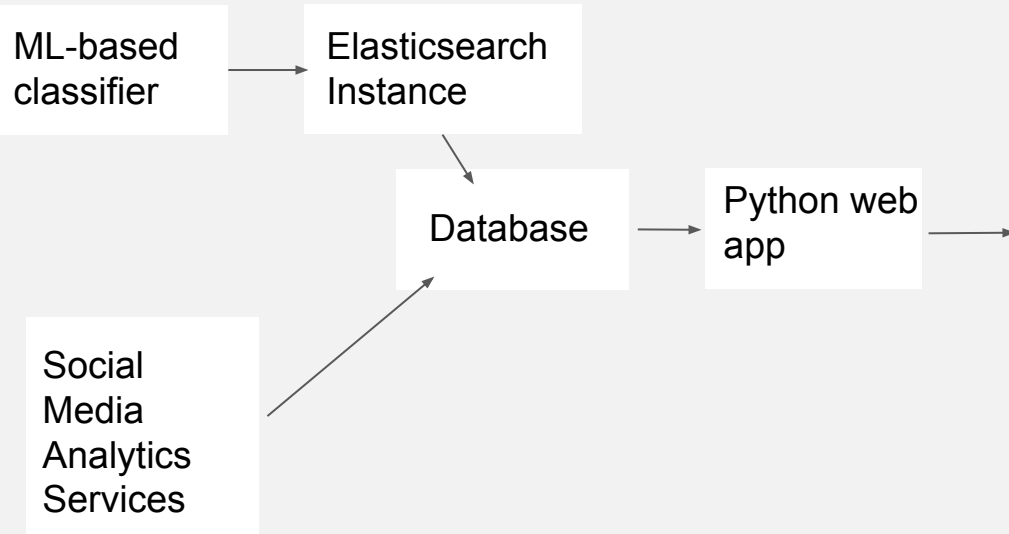
blogs.bournemouth.ac.uk

E.g.

- Packaging/containerising
- Streamlining code - readability, collaboration, usage
- Web apps & APIs
- Measuring software usage & impact
- and more!



Case Study: Impact Enhancement



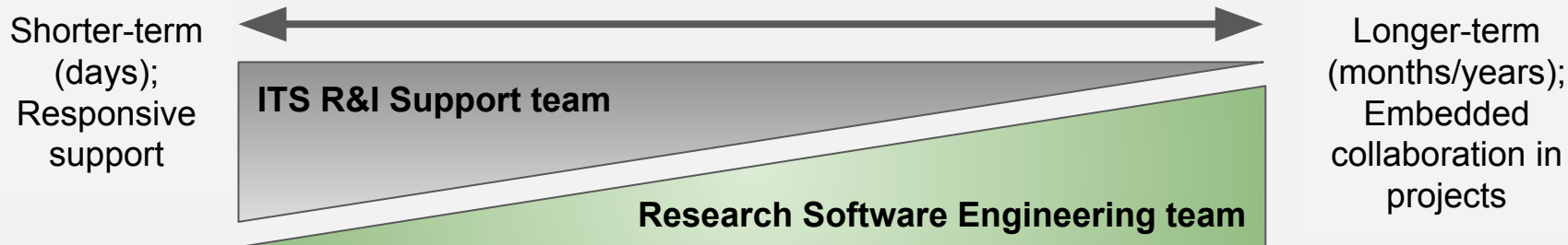
<https://firstdraftnews.org/long-form-article/data-deficits/>





The
University
Of
Sheffield.

Partnerships and sustainability



RSE team not an island: work with ITS R&I, Library, Bioinformatics Core, etc.



The
University
Of
Sheffield.

What we can offer

- Collaboration (Projects)
 - RSEs **working as part of grant-funded projects** (%FTE or per day)
 - **Pre-award consultation required**
 - Some support for post award projects
- Education (collaborative support co-run with ITS R&I)
 - **Training, mentoring and training development**
 - Free Code Clinics: **consultation service**
- Community (for RSEs and Research Developers)
 - LunchBytes: monthly **series of talks** from the research community on research software, data and infrastructure.
 - **Collaborative coding events** e.g. ReproHacks, GPU Hackathon, Hacktoberfest
 - **Mailinglist, Newsletter** and **blog**



The
University
Of
Sheffield.

Getting in touch

rse@sheffield.ac.uk | [@RSE_Sheffield](https://twitter.com/RSE_Sheffield) | <https://rse.shef.ac.uk>

Q&A

Jamboard: <http://bit.ly/RSE2021QA>