

University of Sheffield Software Survey

Robert Turner, Paul Richmond, RSE Team

August 02, 2021

W I P

1 Major Findings

- 382 respondents.
- 91% of respondents use research software (92% nationally).
- 65% report that software is vital to their research (nationally, 69% report that “It would not be practical to conduct my work without software”).
- 27% develop their own code (56% nationally).
- 69% (of the 27% subset who responded to this question) feel they have **not had sufficient training** to develop reliable software.
- Of the 54% of respondents who are involved with writing funding proposals, 45% expected to write software as part of the proposal.
- Of those who expected to write software, 40% did not request funding for this (compared to 20%, nationally).

2 Recommendations

- Investigate why 69% of respondents feel they have not had sufficient training.
- Provide training / support to increase researcher confidence with version control, continuous integration and unit testing.
- Investigate why participants responded as they did to a question about level of support for software development. Create a target for this and monitor performance against it.
- Provide additional support to researchers to make research software outputs not intended for commercialisation freely available.
- Discover if any action can be taken to help researchers would like to use University of Sheffield HPC but don't currently use to do so.
- Advocate for researchers to include costs for software development in their funding applications.

3 Introduction

4 Sample characteristics

This survey and those we make comparisons with used different sampling strategies:

- This survey went to all PhD students and research staff with a prize incentive.
- A 2020 Southampton survey went to “*all staff employed on an ERE contract (Education, Research and Enterprise) and all PhD students*” with a prize incentive. As surveys were sent out on a faculty by faculty basis, it was possible to report a response rate of between 8% and 11% for all faculties.
- 2014 national survey - it is not clear what sampling strategy was used.

In this survey, participants were asked in which faculty they are based in a list of University of Sheffield (UK) faculties and professional services divisions.

We do not know the relative sizes of the different faculties which makes it hard to tell if there is a bias in our responses towards the faculties with higher response rates.

Participants were asked where the funding from their research comes from, and could enter more than one source:

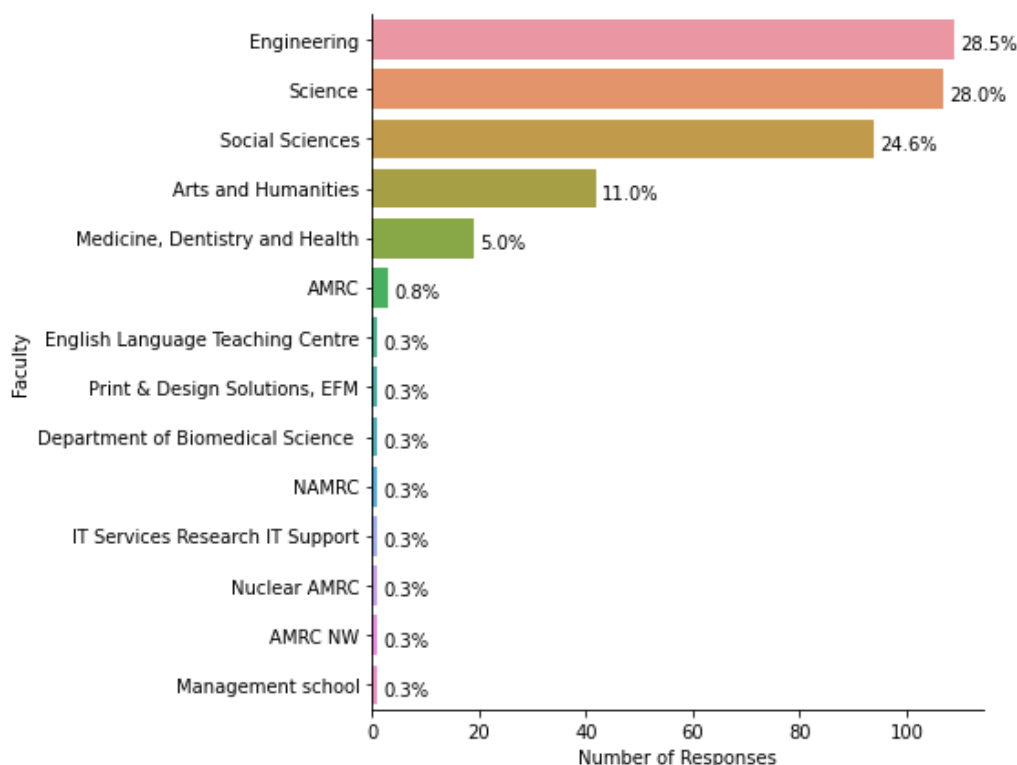


Figure 1: In which faculty are you based?

Where there were 4 or less responses for a given agency, these were grouped as “All others.”

All of the major branches of UKRI are represented, along with charities including the Wellcome Trust and Leverhulme Trust. This chart can’t be interpreted as indicating which bodies are more or less likely to fund software development due to sample bias in which researchers responded to the survey. The largest funder by proportion was the EPSRC, and the largest faculties by response proportion were Engineering and Science, who might be expected to draw most from this funder. We also have not corrected in any way for funder budgets, or number of applications funded. It is not clear that all of these funders funded software development.

Respondents job titles (or student status) show that our sample contained a large proportion of PhD students and Research Associates.

4.1 Optional questions

If someone was a no response for one optional question, it looks like they were also a no response for all the other optional questions.

5 Prevalence and importance of research software

91% of participants report that they use research software (defined as “...any software you have used in the generation of a result that you expect to appear in a publication. This might be anything from a few-line script to clean some data, to a fully fledged software suite. It includes code you have written yourself and code written by someone else.” in the survey form). A 2014 study reports that nationally this is **92%**, another from the University of Southampton in 2020 reports **95%**.

This indicates that research software is critical for researchers at the University of Sheffield, and that this is not exceptional.

This is the vast majority of participants and indicates that research software is near ubiquitous amongst researchers at the University of Sheffield.

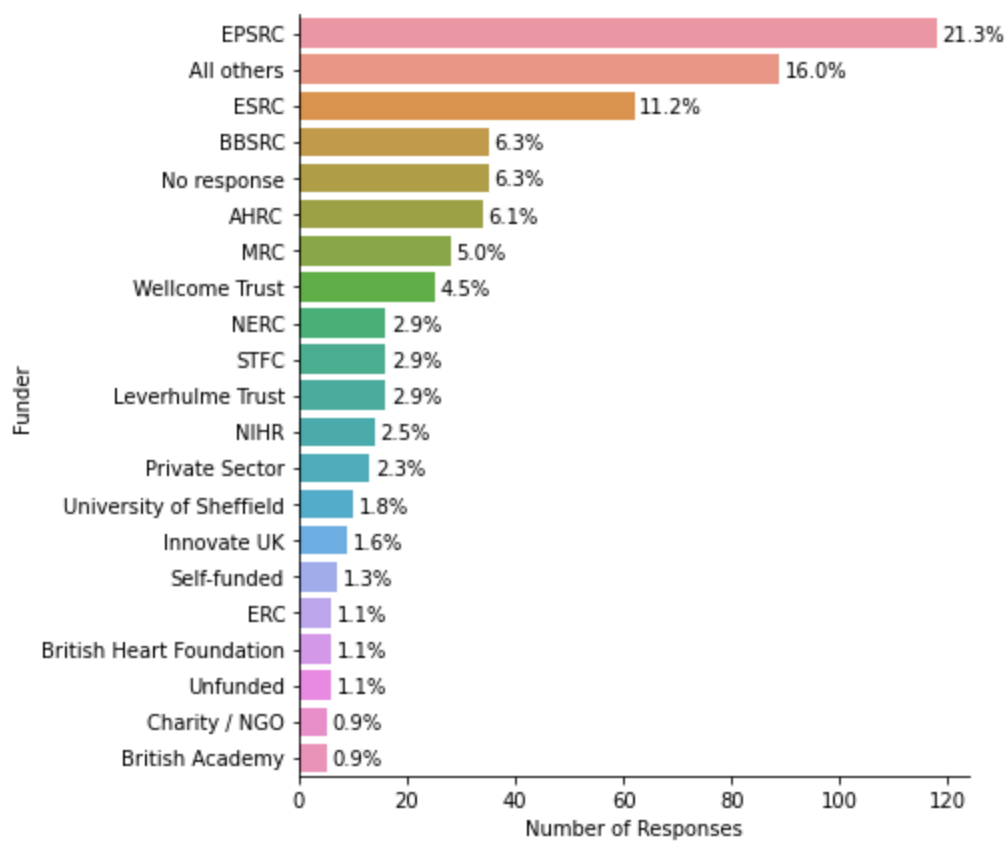


Figure 2: Which of the following organisations usually fund your research?

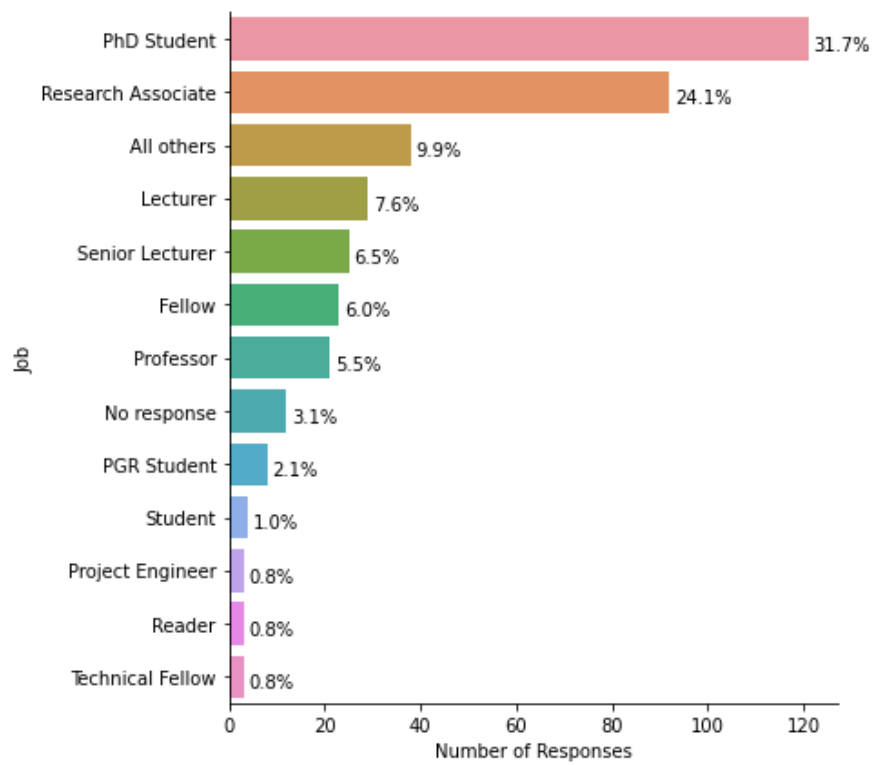


Figure 3: What is your job title?

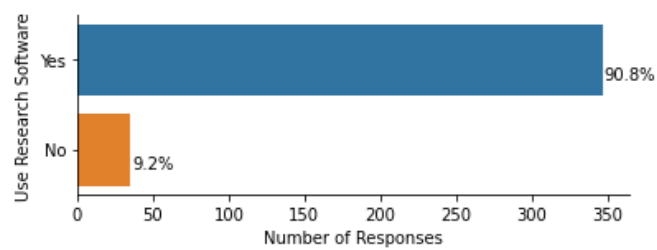


Figure 4: Do you use research software?

Participants were asked to express how important research software is to their work, with 1 being “*Not at all*”, and 5 “*Vital*”.

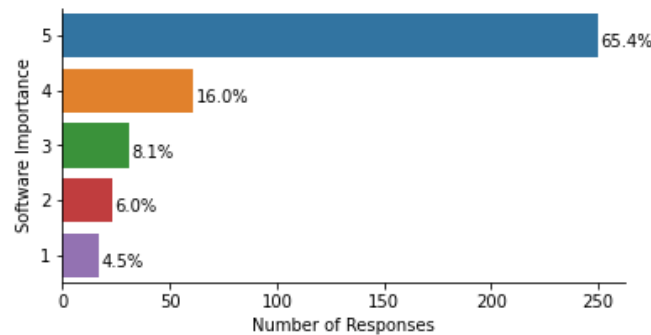


Figure 5: How important is research software to your work?

65% of participants reported that research software is vital to their work.

6 Software development practise

Only 27% of our participants had developed their own code. This compares with 56% nationally. This could be because a smaller proportion of researchers at the University of Sheffield develop code than that for the whole country, or due to sampling bias in either survey. In the 2020 Southampton survey, which records an 8% to 11% response rate accross faculties, suggesting low sample bias, 33% report developing their own software - much closer to our figure. This is consistent with the national survey response having been somewhat biased towards people more engaged with software. Nonetheless, this remains a large proportion of researchers and provokes a question as to whether enough attention is given to software development in our organisation.

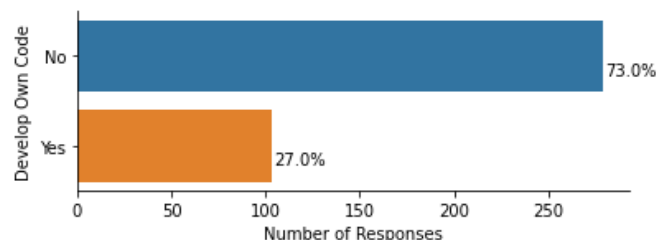


Figure 6: Have you developed your own research software?

The response rate for our question on self-assessment of software development expertise had a low response rate (27% responded). The use of an interval scale introduces challenges in interpretation of results between 1 which is defined as “*Beginner*” and 5, “*Professional*”.

Self-assessment will lead to bias as some people will rate themselves more highly than others. However, we do not see a substantial skew towards *Beginner* or *Professional* - most respondents see themselves as somewhere in the middle.

The question of whether people have the expertise they need to do the software aspects of their research was further examined by asking about sufficiency of training.

Again the number of responses was low (27%), but a majority felt that they had not had sufficient training. This might be because the training cannot be accessed (for a number of reasons including cost, time constraints, or not knowing about it) or is simply not available. This should be investigated further.

We asked later in the survey about confidence in (application of) specific technologies that are important for well engineered software:

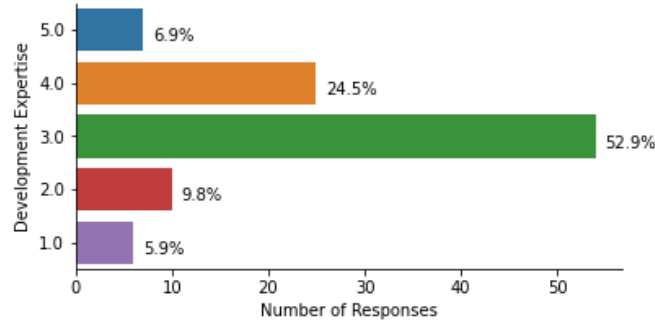


Figure 7: How do you rate your software development expertise?

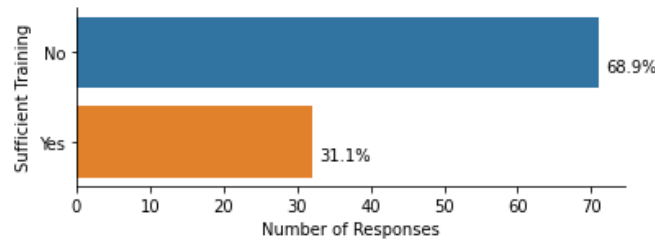


Figure 8: Do you feel that you have received sufficient training to develop reliable software?

- Version control
- Continuous integration
- Unit testing

Here we show results for the (27%) subset of participants that responded to these questions:

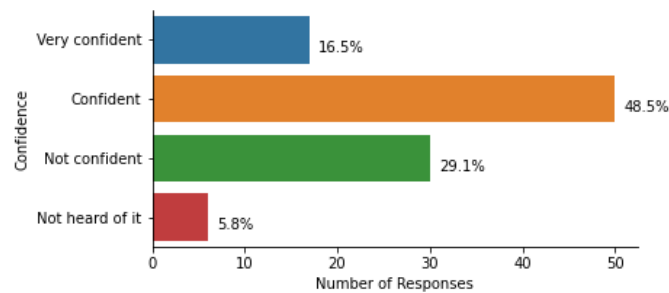


Figure 9: Version control

It appears that in each case either a sizable minority, or a majority of respondents have either not heard of, or are not confident in each of these technologies. Due to their importance for research software engineering, this indicates more should be done to raise confidence.

The question of support for software development was raised from the perspective of what researchers get from “the university”, with a 26% response rate. An interval scale between 1: “*poor*” and 5: “*excellent*” was used.

Clearly, level of support provided depends hugely on resources available. And responses are subjective - we don’t all have the same definition of “poor” and “excellent”. The majority of responses were in the middle of the scale, but at the extremes there were more “poor” (12%) responses than “excellent” (4%). Even the best

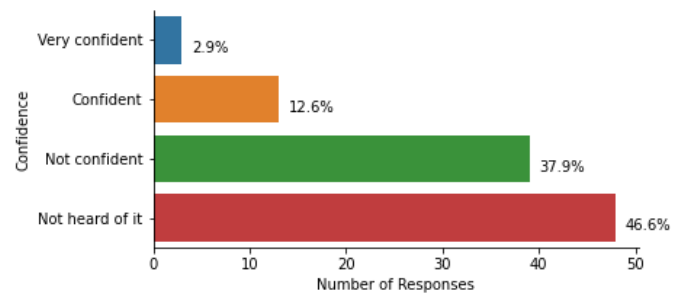


Figure 10: Continuous integration

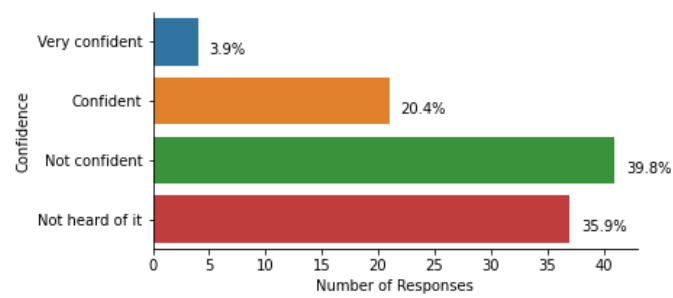


Figure 11: Unit testing

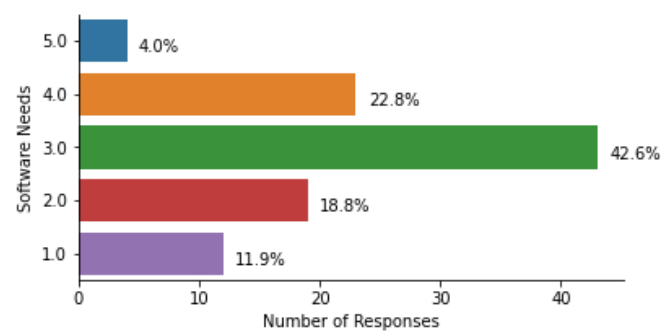


Figure 12: How would you rate the university's current level of support for your software-development needs?

things get poor reviews sometimes - this can be more down to the reviewer than the thing.

7 Funding for Software Development

We found that 44% of participants were involved in writing funding applications.

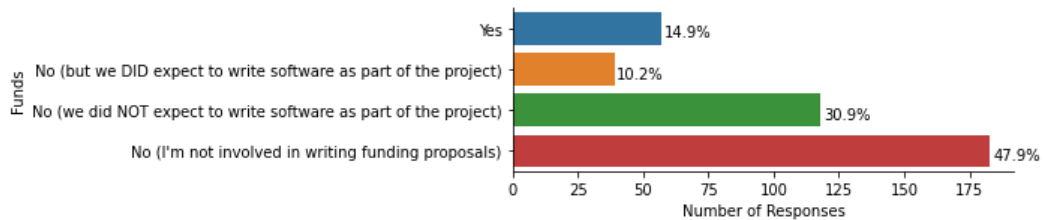


Figure 13: Have you ever included costs for software development in a funding proposal?

If we exclude those respondents not involved in writing funding applications, 45% expected to write software and of these 40% did not request funds for this.

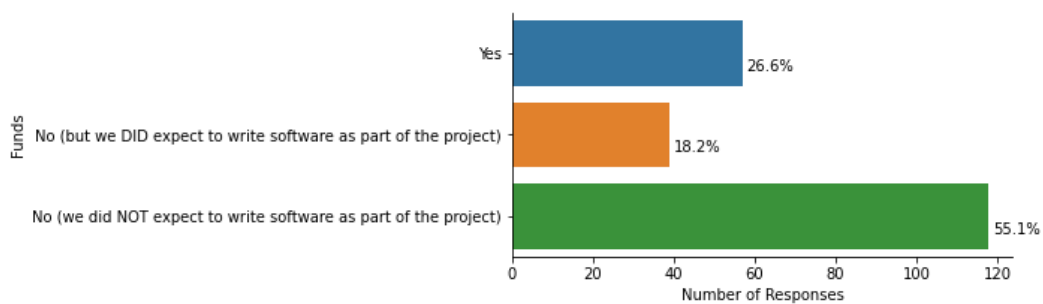


Figure 14: Have you ever included costs for software development in a funding proposal? (excluding those who not involved in writing funding proposals)

8 Staffing of software development

Participants were asked about the suitability of models for staffing their software development needs: “*How suitable would the following models be for your software development needs?*”. These were either hiring a full time developer or using a fraction of an RSE. The response rates for these questions were 89% and 92% respectively.

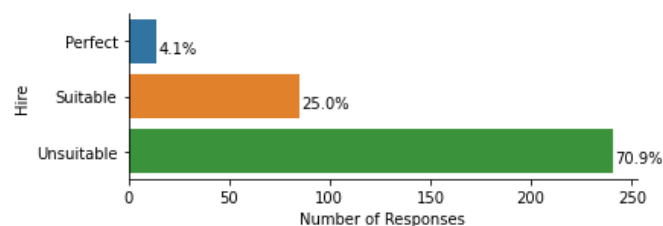


Figure 15: Hire a full-time software developer

In practise, 15% of respondents research teams had hired a developer.

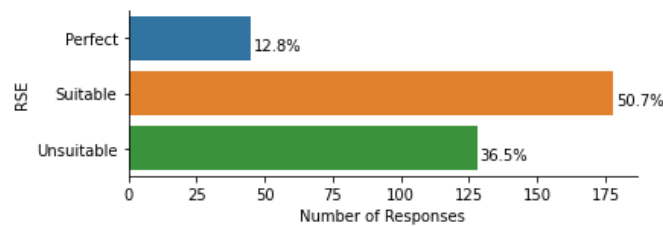


Figure 16: Recruit a developer (or fractional FTE equivalent of a developer) from a central University of Sheffield pool as needed

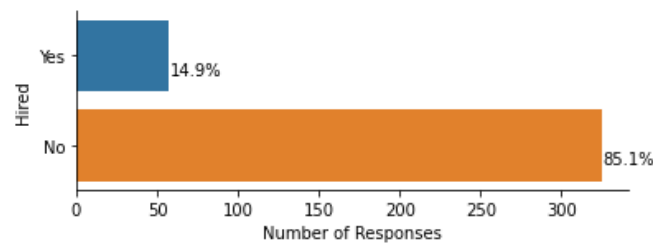


Figure 17: Have you or someone in your group ever hired someone specifically to develop software?

These results indicate that respondents felt it was generally preferable to make use of a pool RSE rather than hire a full time developer.

9 High Performance Computing

Only 27% of participants responded to our question about the use of High Performance Computing (HPC) - most other optional questions had a similarly low response rate, so lack of response here should not be interpreted as lack of interest in HPC. Of those who responded, 25% did not require HPC.

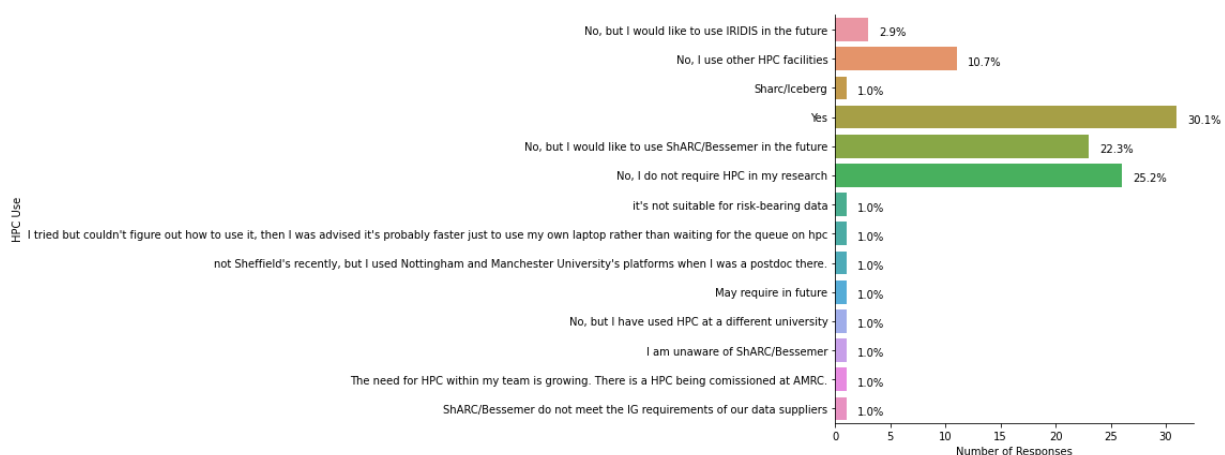


Figure 18: Have you used ShARC/Bessemmer, the University's high-performance computing (HPC) system?

A broad range of possible responses were provided, here. Most important numerically were that 30% of respondents use the University of Sheffield HPC systems, and 22% would like to.

10 Licensing and commercialisation

Response rate for both questions on commercialisation was 27% and results are presented as proportions of those who responded.

A minority of respondents were interested in help from “the university” with commercialising some of their research software.

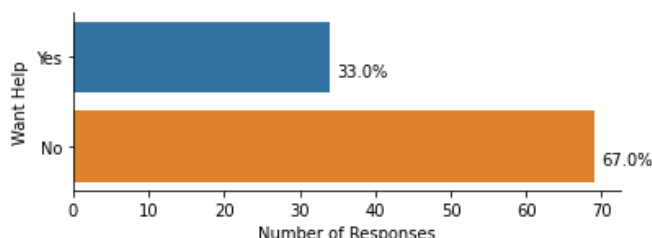


Figure 19: Would you be interested in the university helping you commercialise some of your research software?

Many respondents (45%) were not interested in sharing their software with a commercial partner and a similar proportion (44%) felt that their software was not ready for this.

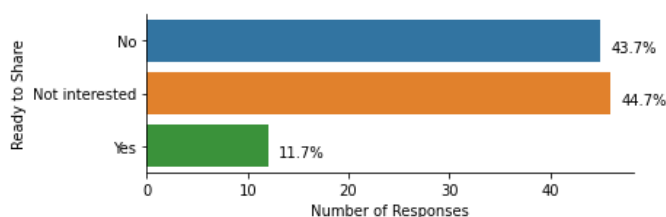


Figure 20: Do you feel that your research software is ready to be shared with a commercial partner?

These results suggest a situation where few participants are interested in or ready to commercialise their software. However, for some, help would be appreciated in doing so.

The University of Sheffield’s Open Research Statement states that researchers should “*Strive to make all scholarly outputs freely available...*” which indicates that software not intended for commercialisation should be made available. This is likely the majority of University of Sheffield research software outputs. Researchers may need additional support to realise this. There may also need to be more support for researchers who are considering, or want to commercialise their software outputs.

11 Methods

11.1 Survey design

The survey was based on previous surveys undertaken nationally and at the University of Southampton.

11.2 Data collection

The survey was sent to all staff and PhD students at the University of Sheffield via email as a linked Google Form. The text of the email is archived in this repository <data/raw/email_announcement.txt>. Responses were incentivised with:

Everyone who completes the survey will have the option to take part in a prize draw for a £50 Amazon voucher and to request a free 1 hour software consultation from our team.

The survey was sent out on 2020-11-09 and closed to responses on ????.

11.3 Anonymisation

11.4 Execution environment

11.5 Data cleaning

Raw data in .csv format underwent modifications to make it suitable for analysis:

- In the fields (job titles, funding sources) where free text fields was allowed, the text entered was mapped to standard responses as much as possible.
- The full text of questions was replaced with a shorter column name.
- Verbose answers from multiple choice were replaced with more concise ones.
- Where respondents could select multiple responses to a question, a separate file was created to accomodate these.
- Where convenient, responses were counted and saved as separate files.

11.6 Plotting

Plotting was done using the **seaborn** Python package in a Jupyter notebook which facilitated iterative improvement of plots.

11.7 Report compilation

This report was written in markdown and converted to .pdf using **pandoc** and MiKTeX.

11.8 Presentation compilation

An accompanying presentation is hosted as a web page built using continuous integration defined with GitHub actions YAML.

12 References

13 Glossary

| Term | Definition |
|---------|--|
| AMRC | Advanced Manufacturing Research Centre |
| AMRC NW | |
| EFM | Estates and facilities management |
| NAMRC | Nuclear Advanced Manufacturing Research Centre |