Track 1 Paper: Surveying the US National Postdoctoral Association Regarding Software Use and Training in Research

Udit Nangia
National Center for Supercomputing Applications (NCSA)
University of Illinois Urbana-Champaign
Urbana, USA
udit.nangia2@gmail.com

Daniel S. Katz
NCSA & CS & ECE & iSchool
University of Illinois Urbana-Champaign
Urbana, USA
d.katz@ieee.org

***Abstract*—This paper reports on the results of a 2017 survey conducted by email and web of members of the US National Postdoctoral Association regarding their use of software in research and their training regarding software development. The responses show that that 95% of respondents use research software. Of all the respondents, 63% state they could not do their research without research software, 31% could do it but with more effort, and 6% would not find a significant difference in their research without research software. In addition, 54% of respondents have not received any training in software development, though all respondents who develop software for researchers have received either self-taught or formal software development training.**

Keywords—software use, software training, survey

I. Introduction

Research software is essential and/or central to a large fraction of modern research projects [1]. Individual researchers choose to use existing software or to develop new software to best carry out their projects. In recent years, computing researchers have explored the relationship between software development and research in academia. For example, a survey of academic faculty and staff at UK universities found that 92% of academics use research software, with 69% saying that their research would not be practical without it. The report also indicated that 56% of respondents develop their own software, and that 21% of them have no training in software development [2, 3]. It is alarming that while a large percentage of research depends on research software and development of this software by academic faculty and staff, a large percentage of software users and a substantial fraction of developers don't have software development training. In addition, a 2008 survey showed that the knowledge required to develop and use scientific software is primarily acquired from peers and through self-study, rather than from formal education and training [4].

This paper expands on the UK data, focusing instead on US universities. It is based on a survey conducted through the membership list of the National Postdoctoral Association, which is an organization that advocates and represents postdoctoral scholars in the US research community. Based on the results of this survey, the paper examines and discusses how role in research, gender, and years of experience influence how researchers utilize research software and the training that they have received in developing research software.

The motivation for the work described in this paper is to assess the role of software in conducting research at US universities. This includes fundamental questions of how prevalent software development training is among researchers and how widespread software use is among researchers at different levels and fields.

II. Methodology

Questions for the survey were based on the questions asked in a similar survey of UK universities [2]. The questions were further modified based on feedback received from conducting a preliminary survey of US universities' faculty and staff. The survey consisted of 14 questions that related to particular aspects of interest in the participants. These included demographic questions, questions assessing the participants' use of research software, and questions evaluating the training that the participants have received in software development.

Respondents who participated in the survey were members of the US National Postdoctoral Association who were present on the organization's mailing list. The survey was distributed through email to the organization's mailing list in March 2017. Members of the organization could input their responses until early April 2017. Members participated purely voluntarily, with no external incentives, and their identities were kept anonymous in collecting the data [5].

Data shown in Table I shows the figures associated with the distribution of the survey. In total, 6281 members of the organization were emailed. Of these, 4473 individuals viewed the email and 209 participants responded to the survey and were included in the analysis. While this is a low response rate (3.33%), it is similar to that of the UK survey [2, 3]. Gizmodo says that response rates can fall below 2% when the respondent population is less-targeted, when contact information is unreliable, or where there is less incentive or little motivation to respond [6]. Because we were unable to provide incentives to respond (other than helping the community) nor were we able to send more than single email request for responses, we are satisfied with this rate, though we recognize there are likely selection effects in our data.

We used descriptive statistical methods to analyze the data obtained from the survey participants. Table II shows the demographic information collected on the participants. Females (55.3%) are more represented than males (42.3%.).

doi: https://doi.org/10.6084/m9.figshare.5328442

And individuals conducting research in STEM fields are more represented in the survey (89.5%) than individuals in non-STEM fields (10.5%). Our survey pool is fairly different than the UK one in terms of gender, with 62% of the UK respondents saying they were male and 36% female.

TABLE I. METRICS ASSOCIATED WITH SURVEY DISTRIBUTION

Total Population Emailed	6281
Total Views	4473
Total Click-Backs	432
Total Responses	209
Response Rate	3.33%

TABLE II. DESCRIPTION OF RESPONDENTS

Gender	***Male***	***Female***	***Other/Prefer not to say***
	42.3%	55.3%	2.4%
Research Discipline	***STEM[a]***	***Non-STEM***	***Other***
	89.5%	10.5%	0%

[a] Science, Technology, Engineering, and Mathematics

III. RESULTS AND DISCUSSION

We are particularly interested in understanding participating researchers' roles, use of research software, and training in software development.

The initial topic in the survey was concerned with assessing the role the respondents play in conducting research. As shown in Fig. 1, 81% of respondents conduct their own research projects; 50% of respondents support someone else's research; 18% manage researchers or research projects, and only 5% of respondents report developing research software for other researchers. This trend of respondents indicating that they conduct their own research having the largest representation and respondents indicating they develop research software having the lowest response is consistent when the results are stratified by different variables, such as gender and years of experience, as shown in Fig. 2 and Fig. 3 respectively. One distinct feature of the data shown in Fig. 2 is that a greater percentage of men report that they develop software for other researchers than women. Fig. 3 shows that a greater percentage of researchers surveyed who had over 10 years of experience develop software for other researchers. This distinction is particularly pronounced because no researchers with 5 or less years of experience reported that they develop software for other researchers.

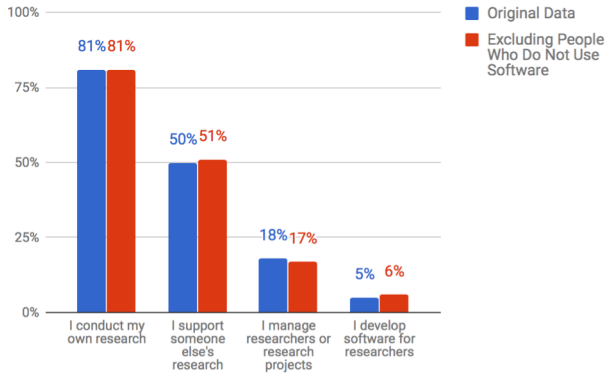

Fig. 1. Assessment of researchers' role between all respondents and respondents excluding those who do not use research software.

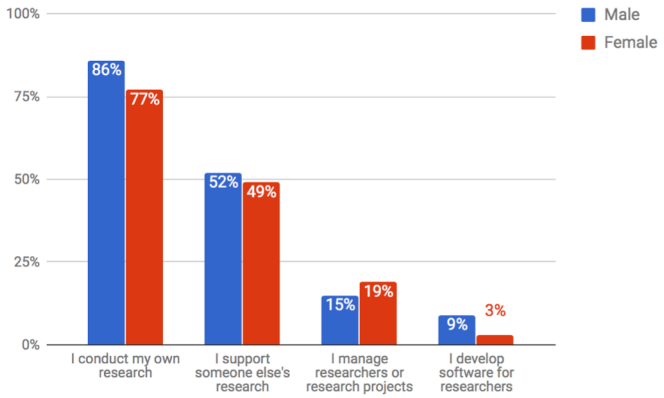

Fig. 2. Assessment of researchers' role between male and female respondents.

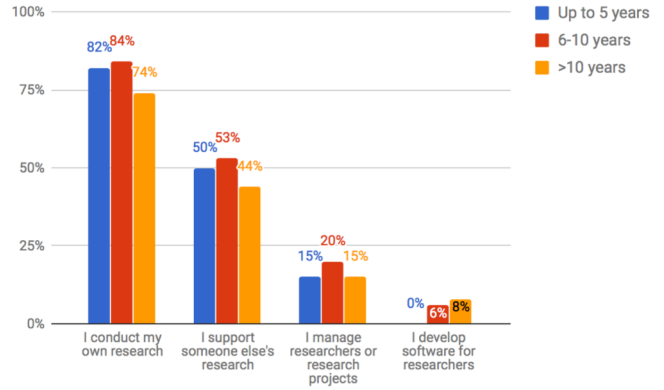

Fig. 3. Assessment of researchers' role between respondents with varying years of experience.

Comparing this study to the UK study, both reported that the respondents conduct their own research at the highest percentage, 81% in the survey of the National Postdoctoral Association and 95% in the survey conducted in the UK Despite this similarity, the results differ greatly, perhaps reflecting the specific groups surveyed. For example, the study conducted in the UK had a much smaller percentage (25%) reporting that they support someone else's research than the survey on the National Postdoctoral Association (50%), while might simply be because postdocs are more likely to support the research of others as opposed to conducting their own research than is the case for university faculty and staff as a whole. The responses received by the surveys also differ in the percentages reported for those who describe their role as managing researchers or research projects. In the survey conducted in the UK, 36% of respondents reported having this role, while in the survey of the National Postdoctoral Association only 18% of the respondents reported that role, which again could simply reflect the role of postdocs. The overall trend also differs between the surveys. In the survey conducted in the UK, more respondents reported that they

manage researchers or research projects than support researchers. This was not true with the survey conducted of the National Postdoctoral Association, which showed the respondents reporting that they support researchers at a greater percentage than they manage researchers or research projects. However, in both surveys the greatest percentage of respondents reported that they conduct research and the lowest percentage develop software for researchers.

The next topic concerned use of research software. As shown in Fig. 4, 95% of respondents say that they use research software. The use of research software remains as a large majority, even when the data is stratified by gender and years of experience, as shown in Fig. 5 and Fig. 6, in which both genders and all levels of experiences report that at least 85% of

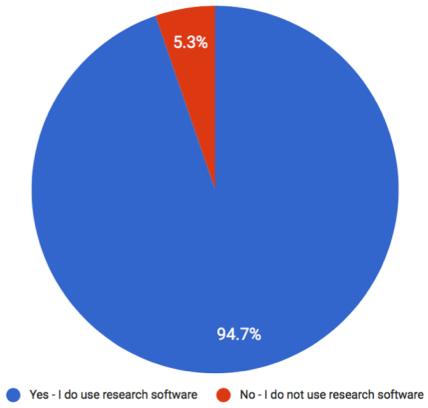

Fig. 4. Use of research software among respondents.

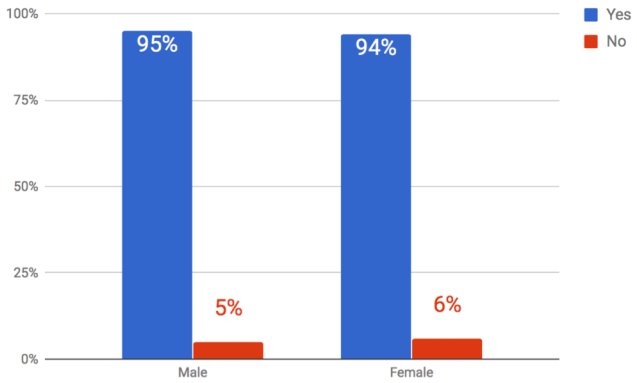

Fig. 5. Use of research software between male and female respondents.

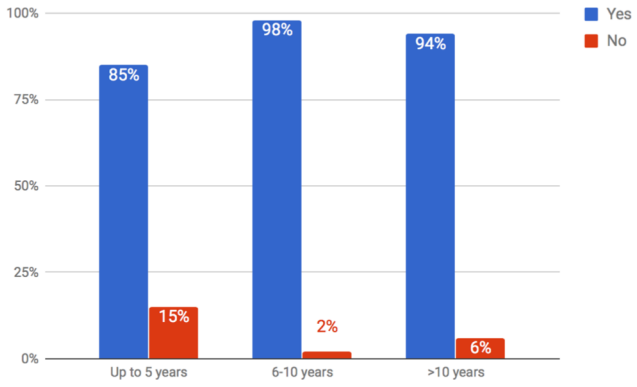

Fig. 6. Use of research software between respondents of varying experience.

respondents use research software. One distinct feature of this data, shown in Fig. 6, is that respondents with only up to 5 years of experience reported that they do not use research software at the highest percentage of 15%. It is also worth mentioning that male and female respondents reported roughly equal percentages of software use as shown in Fig. 5.

In the UK survey, 92% of respondents said they used research software, and this was consistent across genders. Use of research software varied a little with seniority: those having worked in research for 6-10 years reporting the most use (98%) and those having worked for more than 20 years in research reporting the lowest use (86%), with those having worked up to 5 years reporting 91-92% use.

The third topic concerns what would happen if the respondents could no longer use research software. This was intended to assess the need for research software among the respondents. As shown in Fig. 7, a majority of respondents said that it would not be practical for them to conduct their work without software. This trend remained consistent when the data was stratified by gender and years of experience, as shown in Fig. 8 and Fig. 9, in which all categories demonstrated a majority of respondents claiming that it would not be practical to conduct research without software. One distinct feature can be seen in Fig. 9; respondents with only up to 5 years of experience reported that not using software would make no significant difference to their work at a much greater percentage than any of the other groupings, corresponding with the fact that members of this group were the most likely not to use research software. Not surpisingly, Fig. 6 and Fig. 9 show a correspondence between those who could do their research without software and those who don't use software in their research, across all levels of experience and gender, but not an equivalence, which is somewhat puzzling and possibly deserves more detailed follow-up. For example, if 5% of male respondents do not use research software, why would only 2% of male respondents say that not having would make no difference to their work, rather than 5%?

The results found in both the US and UK studies in this area (noting that the questions and possible responses were identical between the surveys) demonstrate the same trend, with respondents stating that not being able to use research software would not be practical to do their work at the greatest percentage, that the work would require more effort but still be possible at the next greater percentage, and that it would make no significant difference at the lowest percentage. In both studies, a majority of respondents claim that it would not be practical to conduct their work without software; this is shown with 63% of respondents in the survey of the National Postdoctoral Association and 69% in the survey conducted in the UK One distinction in the results between the surveys is that in the survey conducted in the UK, a larger percentage (10%) of the participants reported that not being able to use research software would make no significant difference to their work than what was seen in our study (6%). This could be due to the different roles of the surveyed populations (postdocs vs faculty/staff), or it could be due to either the different times in which these surveys were performed, or possibly because of other differences between the US and UK

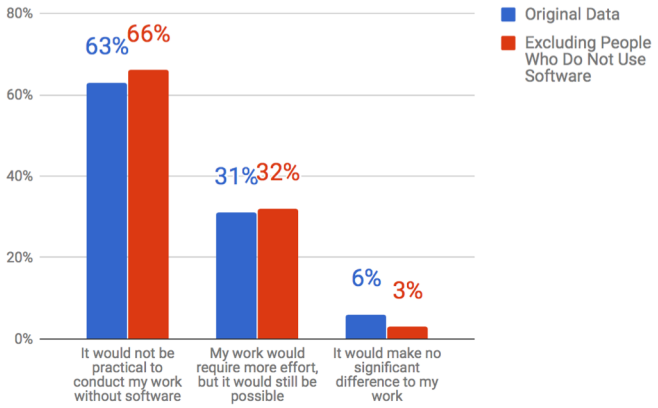

Fig. 7. Assessment of need for research software among all respondents and respondents excluding those who do not use research software.

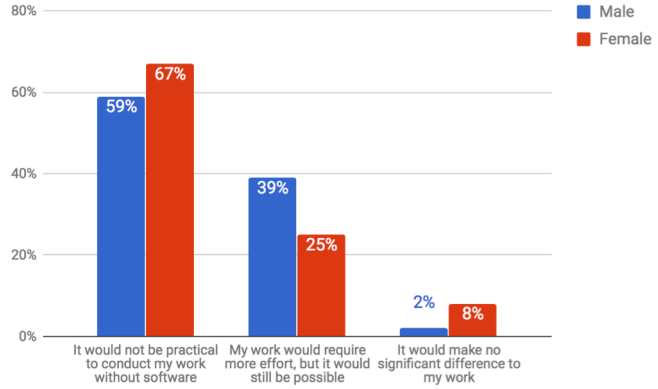

Fig. 8. Assessment of need for research software among male and female respondents.

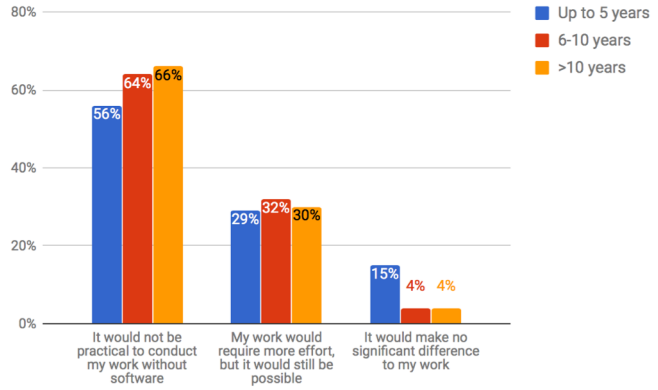

Fig. 9. Assessment of need for research software between respondents of varying experience.

Finally, in order to assess the level of training that the respondents have received in software development, the survey asked "Have you received any training in software development?" The data presented in Fig. 10 summarizes the overall results of this question. The majority of respondents (54%) reported that they had received no training in software development (Fig. 10). Some deviation is seen with regards to gender, as shown in Fig. 11, where female respondents had a large majority (68%) who had not received any training in software development, while male respondents had a majority (64%) reporting that they had received software development

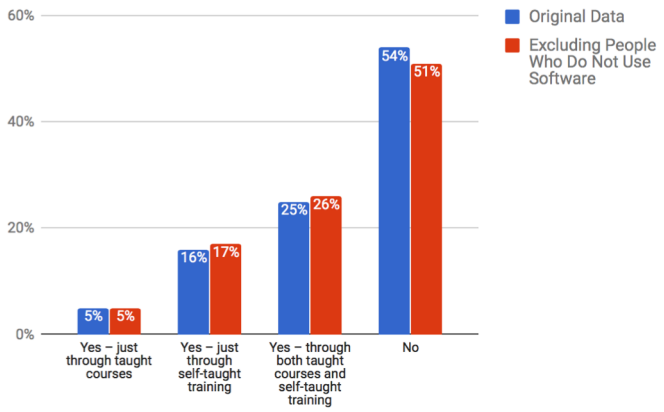

Fig. 10. Training in software development between respondents and respondents excluding those who do not use research software.

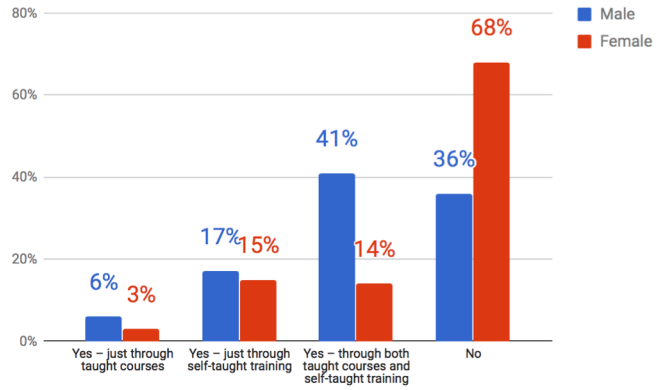

Fig. 11. Training in software development between male and female respondents.

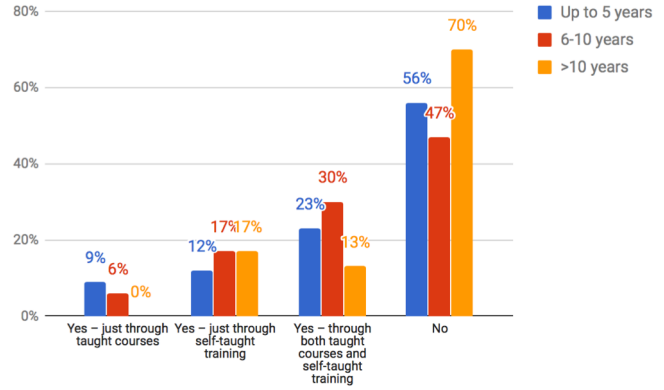

Fig. 12. Training in software development between respondents of varying experience.

training in some form. Fig. 12 shows that respondents with all levels of experiences reported "No" at the greatest percentage. Respondents with more than 10 years of experience not only reported the greatest percentage of individuals that had not received any training in software development (70%) out of all of the groups, but also showed that 0% of the respondents had received training in software development through taught courses.

Comparing the US and UK studies in this area finds that in the UK study, a majority (55%) of the respondents reported

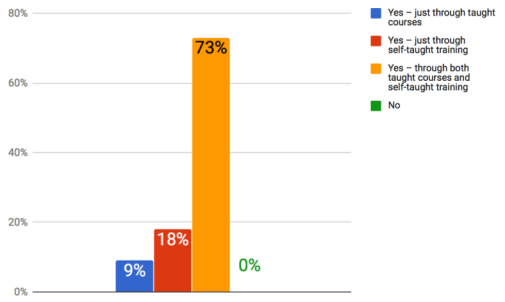

Fig. 13. Training in software development for respondents who develop software for other researchers

that they had received training in software development, but in the US study, a majority (54%) had not received any such training. The results were similar in that in both surveys the lowest percentage of respondents reported that they received training in software through only taught courses; however, this was particularly pronounced in the study in the UK in which no one (0%) reported receiving training through only taught courses.

More positively, training results for the subset of respondents who produce software for other researchers are shown in Fig. 13. All of them have been trained in software development, through self-taught training, instructed courses, or both. And the vast majority (82%) have received training in software development through taught courses.

This is strikingly different from what the UK survey found: only 55% of the UK respondents had received training in software development (15% self taught and 40% had received some form of taught course), and 21% of the UK respondents who develop their own software had no training in software development.

IV. Conclusion

This paper reports on a survey-based study that was conducted to assess the prevalence and need of software use in research across genders and varying years of experience by administering a questionnaire to members of the US National Postdoctoral Association in 2017.

Although the study was conducted at a limited scale, the results of the survey highlight several key issues concerning the use of software in academic research. Research software is in widespread use among respondents, where it plays a vital role in conducting their research. And the small fraction of respondents who develop software for others have had software development training, and most have had formal training.

We recommend that future studies expand on this research by conducting a more comprehensive survey of researchers (beyond postdocs) in the US. We also recommend that future studies work to assess the quality of software produced in research settings.

Although the survey was conducted at a limited scale, its results highlight several key issues concerning the use of software in academic research. Data obtained from the survey showcase the widespread use of research software among respondents and the vital role that software plays in conducting research among respondents. Despite these factors, there is also an evident lack of training in software development present among researchers and a low representation of individuals who produce software for other researchers. While all those surveyed who develop software have some form of training, almost half of the research software users have not. This can be viewed in one of two way: those who do not develop software don't need to be trained in how to do so, or those who have not been trained cannot develop software. Determining if one or both of these are correct will require further study. In general, we believe that future studies should expand on this research by conducting a more comprehensive survey of researchers in the US. We recommend that those future studies work to assess the quality of software produced in research settings. The future studies could ideally be coordinated, so that their data would be easily brought together, both logitisticly and intellectually.

Together, these studies provide evidence that software is imporant in modern research. While more work is needed to examine the rate of change of the importance of software, as well as differences across fields, these studies are starting points for that future research.

Appendix. Questions and Allowed Responses Used in Survey of National Postdoctoral Association.

Question 1: In which institution are you based?
[short text answer]

Question 2: Which of the following roles apply to you?
[selection of one or more of the following responses; if other, short text answer also allowed]

- I conduct my own research
- I support someone else's research
- I manage researchers or research projects
- I develop software for researchers
- Other

Question 3: Which of the following disciplines best applies to your work?
[selection of one of the following responses]

- Chemical sciences
- Biological sciences
- Medical & health sciences
- Physical sciences
- Mathematical sciences
- Earth sciences
- Agriculture, veterinary, environment
- Psychology
- Law
- Philosophy
- Economics
- Human society
- Journalism, library
- Education
- The arts
- Management
- Engineering
- Language
- History
- Politics and policy
- Architecture
- Computing, information science

Question 4: Which funders currently provide the majority of your external funding?
[selection of one or more of the following responses; if other, short text answer also allowed]

- United States Department of Agriculture (USDA)
- Department of Energy (DOE)
- National Science Foundation (NSF)
- National Aeronautics and Space Administration (NASA)
- National Institutes of Health (NIH)
- Department of Justice (DOJ)
- Environmental Protection Agency (EPA)
- Department of Defense (DOD)
- Institute of Education Sciences
- Congressionally Directed Medical Research Programs (CDMRP)
- Advanced Research Projects Agency - Energy (ARPA-E)
- National Endowment for the Arts (NEA)
- National Endowment for the Humanities (NEH)
- Social Science Research Council (SSRC)
- Alfred P. Sloan Foundation
- Gordon and Betty Moore Foundation
- Andrew W. Mellon Foundation
- Bill & Melinda Gates Foundation
- Simons Foundation
- W. M. Keck Foundation
- Don't know
- Other

Question 5: How many years have you worked in research?
[selection of one of the following responses]

- Less than a year
- 1-5 years
- 6-10 years
- 11-15 years
- 15-20 years
- More than 20 years

Question 6: Do you use research software?
[selection of one of the following responses]

- Yes
- No

Question 7: What would happen if you could no longer use research software?
[selection of one of the following responses]

- It would not be practical to conduct my work without software
- My work would require more effort, but it would still be possible
- It would make no significant difference to my work

Question 8: Do you develop your own research software?
[selection of one of the following responses]

- Yes
- No

Question 9: Have you received any training in software development?
[selection of one of the following responses]

- Yes – just through taught courses
- Yes – just through self-taught training
- Yes – through both taught courses and self-taught training
- No

Question 10: Have you ever included costs for software development in a funding proposal?
[selection of one of the following responses]

- I'm not involved in proposal writing
- Yes
- No
- No, but we expected to develop software as part of the project

Question 11: Please provide the name(s) of the main research software you use (up to 5).
[short text answer]

Question 12: What is your job title?
[short text answer]

Question 13: What is your gender?
Respondents could choose one of the following responses:

- Male
- Female
- Other
- Prefer not to say

Question 14: What is your preferred operating system?
[selection of one or more of the following responses; if other, short text answer also allowed]

- Microsoft Windows
- Mac OS X
- Linux
- Don’t have a preference
- Other

REFERENCES

[1] J. Howison and J. Bullard, “Software in the scientific literature: Problems with seeing, finding, and using software mentioned in the biology literature,” J Assn Inf Sci Tec, 67: 2137–2155. 2016. doi: https://doi.org/10.1002/asi.23538

[2] S. Hettrick, “It's impossible to conduct research without software, say 7 out of 10 UK researchers,” Software Sustainaiblity Institute, 2014. Available at: https://www.software.ac.uk/blog/2016-09-12-its-impossible-conduct-research-without-software-say-7-out-10-uk-researchers

[3] S.J. Hettrick, M. Antonioletti, L. Carr, N. Chue Hong, S. Crouch, D. De Roure, et al, “UK Research Software Survey 2014”, Zenodo, 2014. doi: https://doi.org/10.5281/zenodo.14809

[4] J. E. Hannay, C. MacLeod, J. Singer, H. P. Langtangen, D. Pfahl and G. Wilson, "How do scientists develop and use scientific software?," 2009 ICSE Workshop on Software Engineering for Computational Science and Engineering, Vancouver, BC, 2009, pp. 1-8. doi: https://doi.org/10.1109/SECSE.2009.5069155

[5] U. Nangia and D. S. Katz, “Survey of National Postdoctoral Association – Dataset,” [Data set]. 2017. Zenodo. doi: https://doi.org/10.5281/zenodo.843606

[6] Gizmodo, “Survey Response Rates,” 2015. Available from: https://www.surveygizmo.com/survey-blog/survey-response-rates/