Understanding Researchers Practices around Software

Yasmin AlNoamany, University of California, Berkeley
John Borghi, California Digital Library

Introduction

Scientific computation is a core component of modern research. Therefore, software is as important when it comes to building upon existing scholarship. However, when compared to those related to software, practices around data sharing and publications have had more attention. Because research software, which includes both the source code and executables used as part of the research process, presents a significant challenge for efforts aimed at ensuring reproducibility, we conducted a survey examining practices and perceptions related to software. More than 300 participants, representing a variety of research disciplines, participated in our survey answering questions about how they use software.

This poster presents an overview of the study we conducted in addition to the results of surveying researchers across different institutions on software usage.

Methodology

In order to understand researchers’ perceptions, priorities, and practices related to computational reproducibility, we collected responses to an online survey focused on how researchers use, share, and value software. The survey, which was open for responses during the month of February 2017, was distributed using the Qualtrics platform (www.qualtrics.com). The survey was advertised through blog posts, social media, and research-related email lists and listservs.

Our survey instrument consisted of 56 questions (53 multiple choices, 3 open response), divided into four sections. In order, the sections focused on:

- **Demographics**: Included questions related the participant’s research discipline, role, degree, age, institution, and funding sources (7 questions)
- **Characteristics of research software**: Included questions related to how the participants use software and the characteristics of their software (17 questions).
- **Software sharing practices**: Included questions related to how participants make their software available to others (18 questions).
- **How researchers assign value to software**: (14 questions).

Findings

Researchers tend to use open source software more than commercial software.

The purpose of using research software

Here are the categories of reasons researchers use software for: To collect data, To collect and analyze data, To conduct experiments, To develop theories.

The top 10 programming languages used by researchers

<table>
<thead>
<tr>
<th>Language</th>
<th>Python</th>
<th>R</th>
<th>SQL</th>
<th>Java</th>
<th>Matlab</th>
<th>C++</th>
<th>C#</th>
<th>Fortran</th>
<th>Perl</th>
<th>PHP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selection</td>
<td>27%</td>
<td>15%</td>
<td>10%</td>
<td>8%</td>
<td>7%</td>
<td>5%</td>
<td>3%</td>
<td>2%</td>
<td>1%</td>
<td>1%</td>
</tr>
</tbody>
</table>

Overview of Software Practices in Scientific Research

Do you write source code? 100%

- 81.11
- 14.49
- 4.4

Do you use source code written by others? 100%

- 85.72
- 3.31
- 0.98

Do you create executables? 100%

- 53.52
- 44.15
- 2.36

Do you use others’ executables? 100%

- 80.26
- 10.43
- 1.71

Conclusion

We found that researchers consider software to be as important as data. Most researchers do differentiate sharing from preservation, but they need tools and guidance on how to preserve their code. The participants indicated that time and licenses are the main constraints of sharing software.

References