

Supplementary Materials for the Sustainable Open Source Strategies IQP

Produced by Alex Harrigan, Amelia Nishimura, Matt St. Jean, and Thierry De Crespigny

Table of Contents

Appendix 1 (pg. 3): Administrative Policy Brief for the IGB

Appendix 2 (pg. 8): Researcher Flyer for Distribution at the IGB

Appendix 3 (pg. 10): IGB Survey Data

Appendix 4 (pg 13): Digital Humanities Survey Data

Appendix 5 (pg 16): List of Interview by Subject Number and Date Conducted

Appendix 6 (pg 18): List of Team Contributions

Appendix 7 (pg 19): Informed Consent Form

Appendix 1

Administrative Policy Brief

Policy Brief:

Encouraging Good Software Practices and Increased Scientific Collaboration at the IGB

Produced by the Sustainable Open Source Strategies Team from Worcester Polytechnic
Institute



Executive Summary:

Modern research institutions use software both to produce and organize their research, making software skills incredibly useful to researchers. However, many researchers are self-taught, lacking a professional background in software development. Because of this, researchers are often unaware of or lack the time to learn the best practices for publishing work in a well-documented, open-source format once they enter the field. This leads to difficult-to-use code being published alongside research, if it is published at all. By helping researchers improve in these areas, software and hardware that accompany research will become easier to reuse and build on. We are recommending that the IGB adopt policies and procedures that better support those looking to publish their software in open source formats.

Context of Problem:

Publication in open source formats increases both the IGB's ability to collaborate internally amongst researchers as well as the public's access to the research that the IGB produces. Open source is a method of publication and sharing that keeps every piece of a program available to the public for free access, particularly all of the source code. It encourages collaboration on projects by facilitating outside input onto a given work through said accessibility. It also ensures that the software that is created alongside a body of research is just as openly accessible as the research itself. That being said, there are several significant obstacles that stand in the way of effective usage of these publication practices.

At the IGB, many researchers do not have significant software development experience as well as access to tools and resources that would enable them to implement good development practices. A survey conducted among IGB staff showed that nearly half (46%) of respondents were self taught when it came to their software development training. It was also noted in the same survey that the same percentage of respondents were not familiar with version control or documentation software. Documentation and version control are integral parts of the software development process. Interviews we conducted showed that staff at the IGB both had to spend extra time comprehending software due to a lack of documentation and also did not often document their work themselves. It should be noted that, as per interviews conducted with IGB staff, this lack of documentation on the part of IGB staff is both due to the large investment of time effective documentation takes to create as well as the fact that the majority of interviewed individuals did not publish the code they wrote, undercutting the need for documentation. There are no tools for version control (like Git/Github) or documentation that are widely advertised to staff, resulting in less effective management of software projects. In addition to the experience level and access to tools, there also exists a dearth of incentives on an institutional level for researchers to publish their software. Researchers are not awarded specific credit for the software that may be published in non-journal formats, and there is not a specific repository for software to facilitate easy access and collaboration. These programs can be large and complex, but the fact that they aren't given credit for their software if it is not published in a formal journal, regardless of the impact of the work, disincentivizes this publication in the first place. The lack of an effectively used internal repository for smaller programs discourages greater software collaboration and sharing within the IGB. Sharing of code does already occur, but does so on an individual basis.

Policy Alternatives:

The current policy of the IGB only concerns journal publications and leaves a hole for non-conventional software/hardware publications. As it stands, credit is awarded to journal publications using the LOM system on a 0.5 to 5 scale, with the exact awarded points being dependent on the impact factor of the work. This system has no mechanism in place to award points to informally published software, should an individual choose to publish said software or hardware project publically on a service such as Github or GrabCAD. Github and GrabCAD, due to their very large user bases, potentially have very large impact factors when used as a means of publication. There is also no widely used service to facilitate the sharing of code internally within the IGB. Both an internal bitbucket server and a cloud based Github do exist for the purpose of internally sharing code but are severely underused. Interviews with IGB staff confirmed that very few people were aware of the servers' existence and those who were did not use them. We suggest the extension of the LOM system to software and hardware that are informally published publicly, rather than just journal publications in order to address this first issue. We have created fliers and suggest the creation of workshops to teach software tools and good practices. Of the 41 people we surveyed at the IGB, 81% of them knew how to program in R while 54.3% of respondents said they had no familiarity with documentation or version control. Given this fact, we suggest workshops that teach techniques in R as well as documentation and version control practices.. Finally, recommend that either the Bitbucket or Github server should be revamped and advertised to increase their usage and facilitate easier collaboration on code.

Policy Recommendations:

1. Revamp the internal bitbucket server or GitHub service for sharing.
 - i. From our interviews, we identified that while not everyone was eager to publically share their code, they were excited at the idea of an internal sharing server where they could share their code and view the code of their coworkers. Accompanying this change should be an internal awareness campaign on sharing code within the IGB with the git service This could be done through email announcements, the creation of pamphlets and an internal website with quick guides on how to use git services.
2. Offer seminars/workshops on how to use personal Github services and software tools
 - a. From the surveys and interviews that were conducted we have identified that providing workshops on the following topics would be very beneficial to staff and would be well attended and received.
 - b. Documentation Workshop which focuses on the importance of documentation, how you should document your code, and goes over some tools which can make the task easier.
 - i. Roxygen: auto-documentation tools for R scripts (Translates specially formatted comments in source into documentation)
 - ii. Pydocs: python auto-documentation tool
 - c. Version Control Workshop (ft. Github) which discusses why Version Control is extremely helpful to all projects and goes over tools that enable easy version control.
 1. <https://learngitbranching.js.org/>
 - d. R and Python Workshops aimed at helping staff improve their code writing abilities focusing on topics like data processing in R, How to write better algorithms, and debugging and error resolution.
 - e. A Github Workshop encouraging open source publishing practices and going over how to manage sharing your project with the public and how to make and use private projects using Github. This could take advantage of service provided by Github where a specialist from their organization will teach these tools/techniques
3. Open Source Licencing/Journals
 - a. Common open source journals for formal publication.
 - i. The Journal of Open Source Software
 - ii. Journal of Open Hardware
 - b. The IGB should select a number of open source licenses that would be used by research staff:
 - i. Popular Open Source Licences include
 1. Apache License 2.0 (Apache-2.0)
 2. 2-clause BSD license (BSD-2-Clause)
 3. GNU General Public License (GPL)
 4. MIT license (MIT)
 5. A more comprehensive list can be found at:
<https://opensource.org/licenses/category>

4. A LOM System change to include sufficiently noteworthy non-conventional software/hardware publications.
 - a. To make this change there needs to be standards put into place for the software and hardware and software that is published
 - i. All publications should be logged internally with the library along with all relevant metadata
 - ii. Software Standards:
 1. All software publications should follow journal guidelines:
 - a. Be open source
 - b. Be feature complete and designed for maintainable extension.
 2. The software published should be properly documented
 - a. Templates for what constitutes proper documentation can be found at: <https://thegooddocsproject.dev/about.html>
 - iii. Hardware Standards:
 1. Mechanical Projects:
 - a. Files should be saved in a format openable in most CAD software.
 - b. Part description and functions should be written up and provided with the respective files.
 2. Electrical Projects:
 - a. All PCB designs and electrical diagrams should be published alongside the documents
 - iv. Open Access
 1. Alongside publishing in an open source hardware or software journal we suggest that software be published publically using Github for software projects and GrabCAD for Electrical/Hardware projects.
 - v. Judging Impact of a Work
 1. The impact of work of an open publication can be judged using the analytics provided by the Github/GrabCAD on the project use and external contributions to the project. Downloads, project contributions, and comments can all be used to judge the impact of a publication and the amount of points can be scaled accordingly.- 5. Who Will Manage it
 - i. Project management should fall under a number of groups within the IGB but should be coordinated by one person. The IT staff could coordinate with team leads to assess the specific needs of each group and create/revamp the server to meet those needs. For the seminars, volunteers within the igb should be collected to lead and teach workshops. Finally the change to the LOM system would need to be instituted at an administrative level.

Appendix 2

Researcher Flyer

OPTIMIZE YOUR RESEARCH WITH OPEN SOURCE

WHY OPEN SOURCE?

You've probably heard of open source. But did you know that it can help you improve your code and research overall?

Publishing your code or work open source can:

1. Ensure easy access later if needed (who knows when you may need it!)
2. Enable you to keep track of the program's development
3. Ensure the reproducibility of your work by making the software you used easy to find. No one likes having to re-create code.
4. Help resolve errors in your work. The open source community are a helpful bunch
5. Make it easy for others to use and build on your work
6. Encourage others to give feedback quickly. Avoid those long wait times



Appendix 2 (cont.)

Researcher Flyer

PUBLICATION

Here are some tools you can use to publish your software:

1. Internal Bitbucket or Github (neither of which will share your work to the public but just to the IGB)
2. Your own Github
3. Blogging
4. Open Source Journals for larger projects such as The Journal of Open Source Software and Journal of Open Hardware (for those of you who do hardware work as well)

DOCUMENTATION

Documentation of your work is key to both your own future understanding of what you did and enabling someone else to build on your work

1. Use Doxygen or Roxygen (if you use R) to auto generate documentation based on in-line comments
2. Truly comprehensive documentation is something that is inherently time intensive but is a critical part of ensuring that your code is accessible for those who want to build on it or even yourself a year or two down the line. This will save you time in the future

VERSION CONTROL

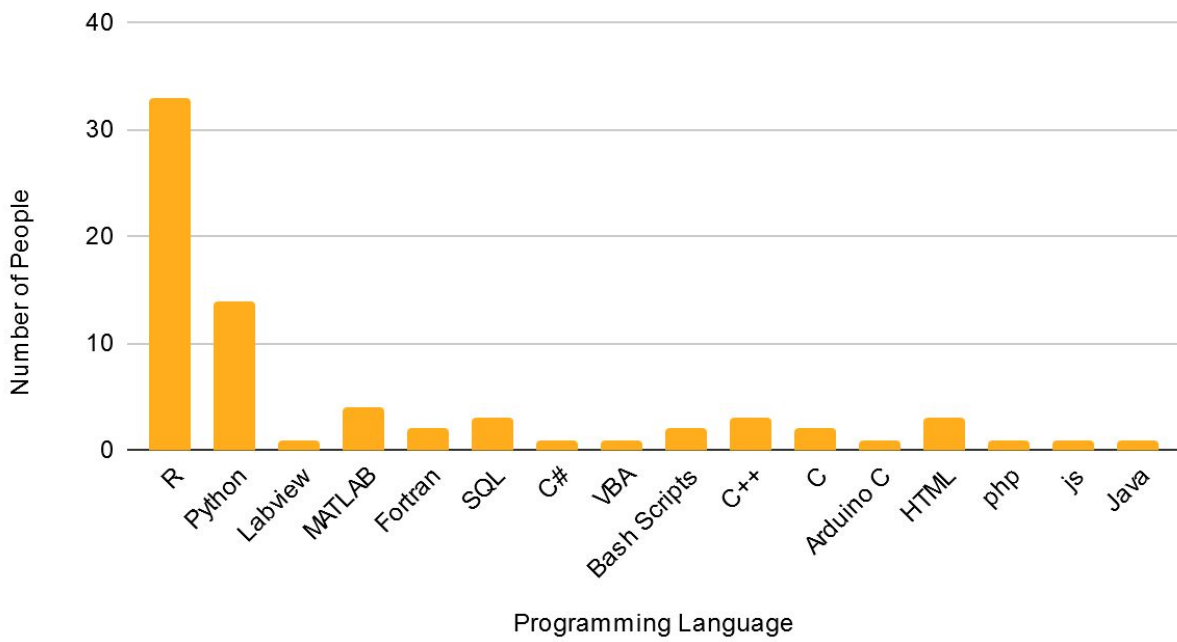
Use Git to manage the updates of your code. Using Github inherently uses Git to push forward changes to the master version

1. Git is a version controls system that uses commands to make changes to a work.
2. It can help you manage branches and forks of your work and integrate changes made by others
3. Using Github also allows you to share your work with the public should you so choose.

Appendix 3

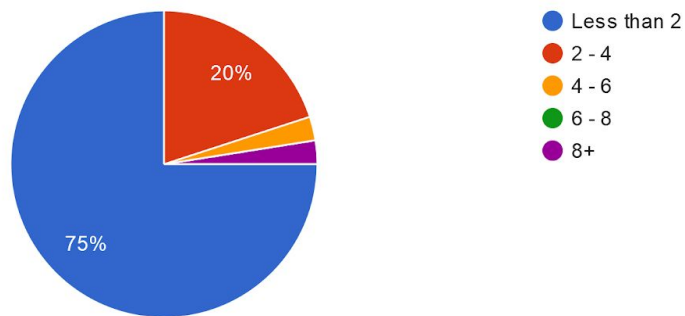
IGB Survey Data

Primary Programming Languages



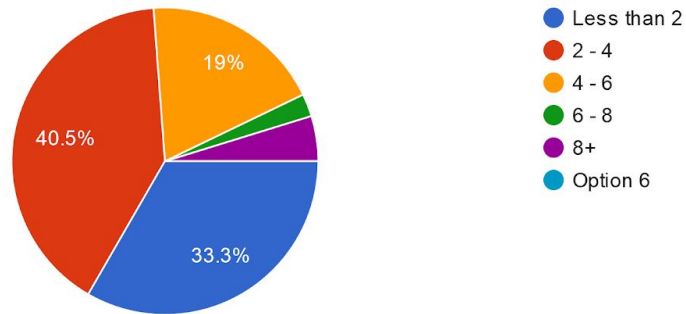
How many hours per day do you spend on average developing scientific software?

40 responses



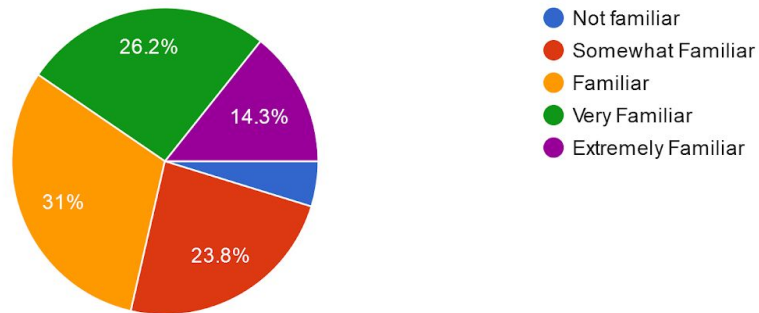
How many hours per day do you spend using scientific software developed by others?

42 responses



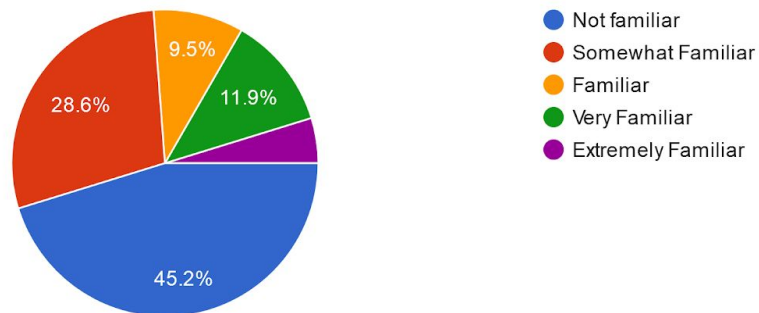
How would you rate your familiarity with the concept of open source?

42 responses



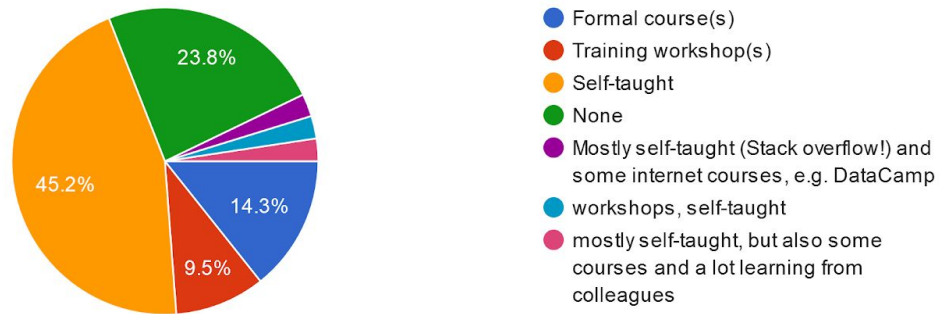
How would you rate your familiarity with version control/documentation software?

42 responses



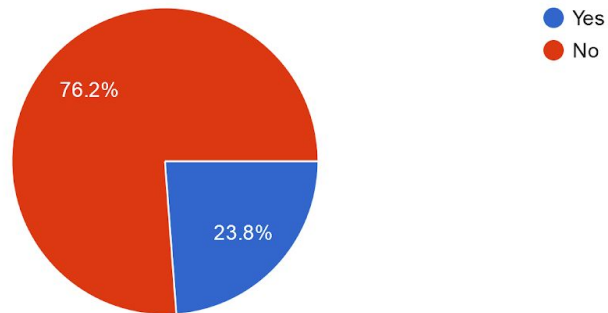
What sort of (if any) software engineering training have you received?

42 responses



Would you be willing to be interviewed as a part of a study on open source in research?

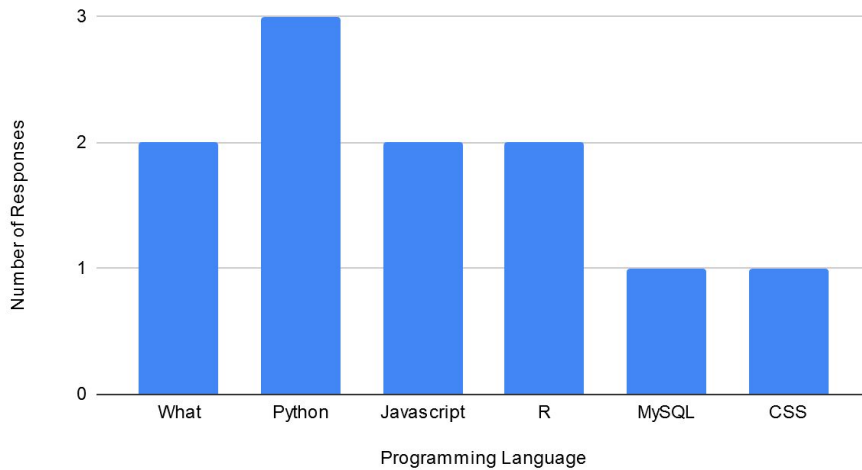
42 responses



Appendix 4:

Digital Humanities Survey Data

What programming languages (if any) do you program in?

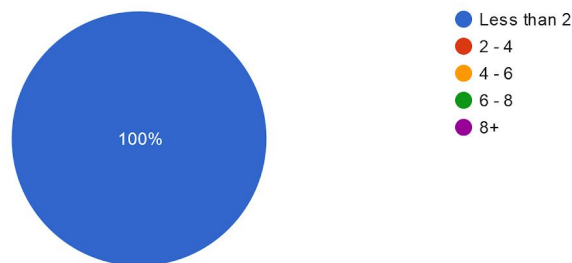


What are the primary tech tools or platforms that you use for your projects (e.g. Omeka, Wordpress, Storymaps, etc.)?

- Response 1: Wordpress, Storymaps
- Omeka, Jekyll, Wordpress
- MALLETT, D3
- HathiTrust, Wordpress, Storymaps (ArcGIS), Voyant
- Omeka, Curatescape, Wordpress, SCALAR, ARCGis, Storymaps
- RStudio

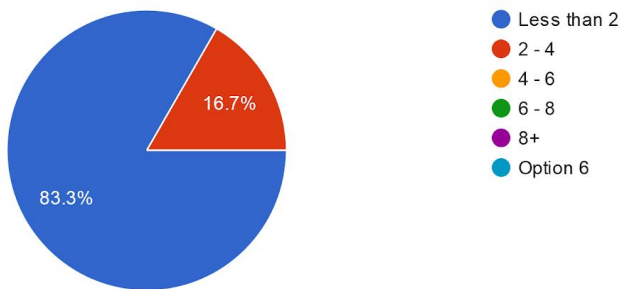
How many hours per day do you spend on average developing software or coding for projects?

6 responses



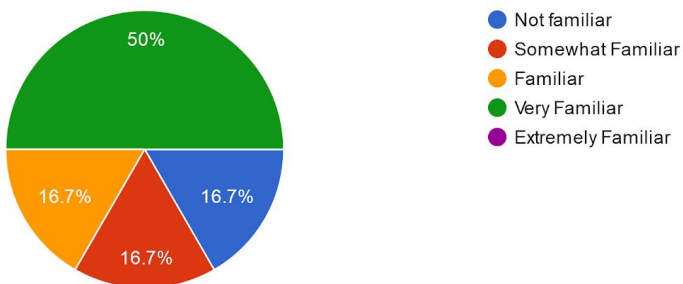
How many hours per day do you spend using software or scripts developed by others?

6 responses



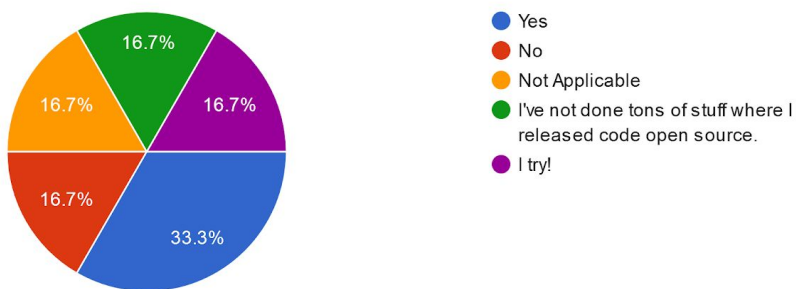
How would you rate your familiarity with version control/documentation software?

6 responses



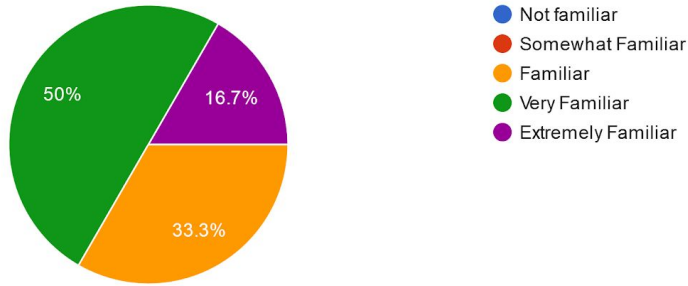
Do you provide documentation with your Digital Humanities projects?

6 responses



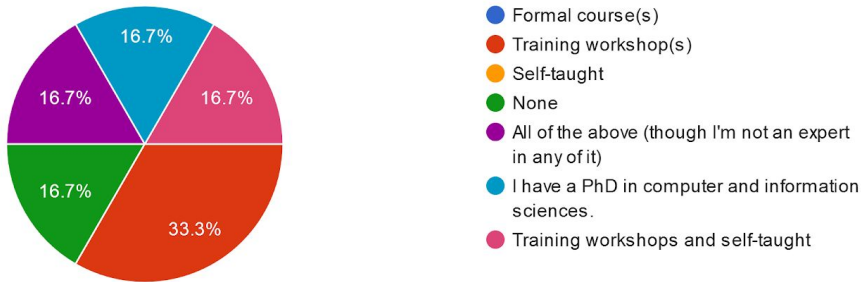
How would you rate your familiarity with the concept of open source?

6 responses



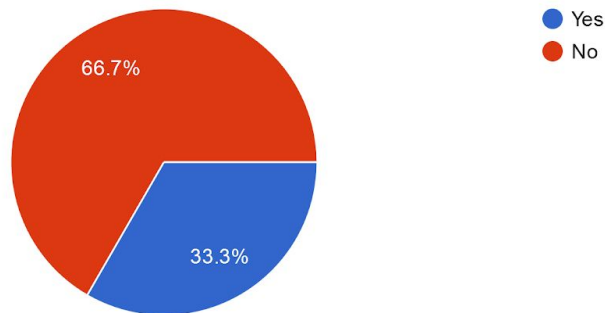
What sort of (if any) software engineering training have you received?

6 responses



Would you be willing to be interviewed as a part of a study on open source in research?

6 responses



Appendix 5

List of Interviews by Subject Number and Date Conducted

Participant Number	Participant Position	Subject Group	Date of Interview
1	PostDoc Scientist	IGB	April 7, 2020
2	Doctoral Student	IGB	April 17, 2020
3	Doctoral Student	IGB	April 17, 2020
4	Doctoral Student	IGB	April 20, 2020
5	Technician	IGB	April 23, 2020
6	Modeler	IGB	April 24, 2020
7	Engineer	IGB	April 27, 2020
8	Research Assistant	IGB	April 27, 2020
9	Head of Student Programs	HUA	April 30, 2020
10	Professor	N/A	April 29, 2020
11	Assistant Professor	HUA	May 1, 2020
12	Research Librarian	IGB	April 15, 2020

Appendix 6

List of Team Contributions

Component	Primary Contributor	Secondary Contributor
Booklet Graphic Design	Amelia Nishimura	
Booklet Abstract	Writing: Matt St. Jean Editor: Matt St. Jean	
Booklet Introduction	Writer: Alex Harrigan Editor: Amelia Nishimura	
Booklet Background	Writer: Alex Harrigan Editor: Amelia Nishimura	Background Info: Full Team
Booklet Methods	Writer: Alex Harrigan Editor: Amelia Nishimura	Editor: Alex Harrigan
Booklet Results/Analysis	Writer: Matt St. Jean Editor: Alex Harrigan	Editor: Amelia Nishimura
Booklet Conclusions/Recommendations	Writer: Alex Harrigan Editor: Alex Harrigan	
Data Formatting for Booklet	Thierry De Crespigny	
IGB Policy Brief	Writer: Thierry De Crespigny Editor: Thierry De Crespigny	Editor: Alex Harrigan
Researcher Flyer	Amelia Nishimura	Text of Doc: Alex Harrigan
A3 Reports	Alex Harrigan	
Sponsor Communications	Alex Harrigan	
Interview Coordination	Alex Harrigan	
Agendas	Matt St. Jean	
Meeting Notetaker	Amelia Nishimura	
Software Research	Thierry De Crespigny	
Slides	Matt St Jean	Thierry De Crespigny, Alex Harrigan
Data Formatting for Slides:	Matt St. Jean	

Conduction of Interviews	All Team Members	
Interview Transcription	Matt St. Jean	Thierry De Crespigny
Interview Notetaker	All Team Members	
Guiding Interview Questions	Matt St. Jean	
Survey Questions	Alex Harrigan	Matt St Jean
Mandatory Fun	Full Team	
Task Management	Alex Harrigan	Matt St Jean
Meeting Lead	Alex Harrigan	
Infographic	Amelia Nishimura	Text Writing: Alex Harrigan
IRB Application	Thierry De Crespigny	

Appendix 7:

Informed Consent Form

Informed Consent Agreement for Participation in a Research Study

Investigator:

Principal Investigator: Katherine Foo

Co-Investigator: Sarah Stanlick

Student Investigators: Amelia Nishimura, Alex Harrigan, Thierry de Crespigny, Matthew St Jean.

Contact Information:

Kfoo@wpi.edu

Sstanlick@wpi.edu

gr-sustainable-open-source@wpi.edu

Title of Research Study:

Sustainable Open Source Strategies

Sponsor:

Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB)

Introduction: You are being asked to participate in a research study. Before you agree, however, you must be fully informed about the purpose of the study, the procedures to be followed, and any benefits, risks or discomfort that you may experience as a result of your participation. This form presents information about the study so that you may make a fully informed decision regarding your participation.

Purpose of the study:

A study on how software can improve research workflows and accessibility.

Procedures to be followed:

Ask for permission to record the interview, proceed with the interview which will take place over zoom, Each interview should take between 20 - 40 minutes and the interview will be used to inform out policy suggestions to the IGB. The audio transcript will be reviewed after the fact to clarify any gaps in the interview notes.

Record keeping and confidentiality:

Information is stored on a private google drive and access is given to members of the team. Information is expected to be reported with or without names and the resultant data is restricted to team members only.

“Records of your participation in this study will be held confidential so far as permitted by law. However, the study investigators, the sponsor or its designee and, under certain circumstances, the Worcester Polytechnic Institute Institutional Review Board (WPI IRB) will be able to inspect and have access to confidential data that identify you by name. Any publication or presentation of the data will not identify you.”

For more information about this research or about the rights of research participants, or in case of research-related injury, contact:

Principal Investigator: Katherine Foo, Email: Kfoo@wpi.edu

Co-Investigator: Sarah Stanlick, Email: Sstanlick@wpi.edu

Student Investigators: Amelia Nishimura, Alex Harrigan, Thierry de Crespigny, Matthew St Jean., Email: gr-sustainable-open-source@wpi.edu

IRB Manager: Ruth McKeogh, Tel. 508 831- 6699, Email: irb@wpi.edu

Human Protection Administrator: Gabriel Johnson, Tel. 508-831-4989, Email: gjohnson@wpi.edu

Your participation in this research is voluntary. Your refusal to participate will not result in any penalty to you or any loss of benefits to which you may otherwise be entitled. You may decide to stop participating in the research at any time without penalty or loss of other benefits. The project investigators retain the right to cancel or postpone the experimental procedures at any time they see fit.

By signing below, you acknowledge that you have been informed about and consent to be a participant in the study described above. Make sure that your questions are answered to your satisfaction before signing. You are entitled to retain a copy of this consent agreement.

_____ Date: _____ Study Participant Signature

_____ Study Participant Name (Please print)

_____ Date: _____ Signature of
Person who explained this study