ITCR Scientific Software Impact and Engagement Analysis Survey

The purpose of this survey is to:

- help us to identify how often ITCR scientific software/tool developers evaluate user engagement of their scientific software
- help us to better understand what motivates people to perform analyses of their software use and impact, what challenges or barriers people face in assessing software engagement, what activities are typically done or not done to support such analyses, and what metrics people use and have found to be helpful
- help us get results for a manuscript about methods for evaluating the engagement of scientific software to provide evidence for challenges on this topic and to discern how often such software assessments occur
- help us identify ways for the ITN to better support the ITCR software developers to perform assessments of user engagement with their software

The raw responses will be evaluated by Carrie Wright and a graduate student Awan Afiaz. Trends and patterns in the data will be further examined by looking at summarized data by the ITCR OPEN (formerly called TOW) participants.

We define **scientific software tools** loosely according to biotoolsSchema, which includes:

- **Computing Web-based Platform** - A website providing computing resources and possibly data
- **Web-based tool** - A tool that runs in your web browser
- **Bioconductor R packages**
- **Other R packages (not Bioconductor)**
- **Jupyter Notebooks**
- **Desktop Application** - A tool that runs on your desktop environment with a GUI
- **Database/Ontology**
- **Plug-in** - A software component encapsulating a set of related functions, which are not standalone, i.e. depend upon other software for its use, e.g. a Javascript widget, or a plug-in, extension add-on etc. that extends the function of some existing to
- **Command-line tool/Other scripts** - A tool that works with a command-line interface or environment
- **Suite** - multiple tools that work together

To ensure the privacy of the participants:

- All responses will be anonymous
- The raw data will only be seen by Carrie Wright (and a graduate student Awan Afiaz)

- Details that could identify a research group will not be shared with funding administrators
- If published or shared later - will only be done so in a summarized form - unless it is an anonymous short answer

Passwords for the gmail account associated with this form will be kept up-to-date and access to the responses will be restricted to only Carrie and Awan.

Please contact Carrie Wright at cwright2@fredhutch.org with any questions or concerns.

You may also contact the Fred Hutchinson Cancer Center IRB Office if you have questions about your rights as a participant/parent of a study participant. Contact the IRB if you feel you have not been treated fairly or if you have other concerns.

This study has been deemed exempt by the Fred Hutchinson Cancer Center IRB.
IRB RG No: 11082
Date Approved: 2022-11-29

The IRB contact information is:
Address: Fred Hutchinson Cancer Center, Institutional Review Office,
1100 Fairview Ave. N., Mail Stop J2-100, Seattle, WA 98109
Telephone: 206.667.5900
E-mail: IRBinbox@fredhutch.org

All questions are optional (except the initial consent questions). Participation is voluntary. The survey should only take roughly 10 min.

Thank you for your participation!

* Indicates required question

1. Are you 18 years of age or older? *

Mark only one oval.

- ◯ Yes
- ◯ No

2. I consent to participate in the survey and understand that my participation is voluntary. *

Mark only one oval.

- ◯ Yes
- ◯ No

Tell us about yourself!

In this first section, we want to ask you a few questions about your involvement with scientific software development.

3. This survey is intended for those who are involved with a scientific **software-related** project **funded by ITCR**. Are you involved in such a project?

Mark only one oval.

- ◯ Yes (continue with the survey)
- ◯ No (finish the survey)

4. What is your current **role** on the tool development/maintenance projects (choose all that apply)?

Check all that apply.

- ☐ Manager/advisor
- ☐ Software Developer/Maintainer
- ☐ Outreach Specialist
- ☐ Trainee - postdoc
- ☐ Trainee - graduate student
- ☐ Other: _____

5. How many projects related to developing scientific software tools have you been involved in?

Mark only one oval.

- ◯ 1
- ◯ 2-4
- ◯ 5-9
- ◯ 10 or more

Previous Evaluation Experience

6. What would be **your goals** in evaluating the impact, engagement, or usage of a software tool? (Note most of the survey is multiple choice! Simple phrases are fine!)

_____

Tell us about the scientific software/tool that is the **most developed/mature** that you have worked on.

Please fill out this portion of the survey with only the **single** most developed/mature scientific software that you have worked on in mind.

7. What type of scientific software/tool did you work on? Please answer for the ***single*** **most developed/mature** tool. (This is loosely based on biotoolsSchema - check all that apply if your tool is built form multiple components, otherwise choose the best option)

Check all that apply.

- [ ] Computing Web-based Platform - A website providing computing resources and possibly data
- [ ] Web-based tool - A tool that runs in your web browser but doesn't necessarily provide access to data
- [ ] Bioconductor R packages
- [ ] Other R packages (not Bioconductor)
- [ ] Jupyter Notebooks
- [ ] Desktop Application - A tool that runs on your desktop environment with a GUI
- [ ] Database/Ontology
- [ ] Plug-in - A software component encapsulating a set of related functions, which are not standalone, i.e. depend upon other software for its use, e.g. a Javascript widget, or a plug-in, extension add-on etc. that extends the function of some existing to
- [ ] Command-line tool/Other scripts - A tool that works with a command-line interface or environment
- [ ] Suite - multiple tools that work together
- [ ] Not Sure
- [ ] Other: _____

8. If you feel comfortable, please provide a **link** to your tool here.

_____

9. What **types of users** might adopt your scientific software tool?

Check all that apply.

- [ ] Institutions
- [ ] Individual researchers
- [ ] Patients or other clinical population such as caregivers
- [ ] Clinicians
- [ ] Other: _____

10. Which of the following have you created in terms of **a web presence** for your scientific software/tool?

Check all that apply.

- [ ] Public code repository like GitHub with a readme explaining what the tool is
- [ ] A separate website for the tool with more information than the code repository readme
- [ ] Created a video describing the tool (not how to use it, just what it does)
- [ ] Enrolled in a review system for users to review the software like https://sourceforge.net/
- [ ] None
- [ ] Other: _____

11. Which of the following **communication strategies** have you implemented or supported to connect with users of your most **fully developed/mature** tool? (select all that apply)

Check all that apply.

- [ ] Workshops / Live sessions
- [ ] Talk/poster at conference
- [ ] Discussion groups: e.g. stack overflow, biostars, quora, Discourse, other
- [ ] Google group or something similar
- [ ] Slack community or something similar
- [ ] Email newsletter
- [ ] Social media presence: twitter, instagram, LinkedIn, ResearchGate
- [ ] None
- [ ] Other: _____

12. Which of the following do you provide for **contact information** to help users use your scientific software/tool?

Check all that apply.

- [ ] Simple contact method for users to email the developers
- [ ] More extensive contact methods for users to report bugs, request help, or otherwise engage with the developers (for example, an issue template on GitHub, a google form etc.)
- [ ] None
- [ ] Other: _____

13. What type of **documentation/training** for users to learn how to use a tool did/do you provide?

Check all that apply.

- [ ] None
- [ ] README file
- [ ] training is built in to the software
- [ ] videos
- [ ] web documentation
- [ ] Book
- [ ] Course
- [ ] Journal publication

14. Which of the following **software health infrastructure** have you implemented for your scientific software/tool?

Check all that apply.

- [ ] version control without automated deployment or delivery
- [ ] version control with automated deployment or delivery (rendering or new code version release)
- [ ] provided users with information about the active number of contributors
- [ ] provided users with license about code reuse
- [ ] provided users with metrics about testing code coverage
- [ ] provided users with a metrics on commit frequency
- [ ] automated testing (unit testing or otherwise)
- [ ] Other automations
- [ ] Other: _____

15. **How** should users **cite** your software?

Check all that apply.

- [ ] You have a specific publication you ask people to cite
- [ ] You use a Digital Object Identifier (DOI)/citation enabler for the software itself (using options like https://zenodo.org/)
- [ ] You provide information to users about how to cite your software
- [ ] You provide information to users about when to cite your software
- [ ] You think users know how to cite your software without explicit instruction
- [ ] You think users know when to cite your software without explicit instruction

16. Have you or your team attempted to **recruit additional users** for the tool (or are you planning to)?

Mark only one oval.

- ◯ Yes, we already have
- ◯ Yes, in the future
- ◯ No
- ◯ Other: _____

17. How would you **classify** your scientific software/tool?

Check all that apply.

- [ ] Omics - proteomics, genomics, metabolomics
- [ ] Clinical
- [ ] Imaging
- [ ] Supports multiple types of data
- [ ] Other: _____

Clinical Impact Metrics

18. Which of the following **metrics** have you used to evaluate the **clinical impact** of your scientific software/tool?

Check all that apply.

- [ ] Adoption metrics - Number or proportion of institutions that implement your tool (if your software is implemented by hospitals, or centers)
- [ ] Patient reach metrics - Number or representativeness of eligible patients whose care is impacted by the tool
- [ ] Patient impact metrics - Number of patients whose care or treatment or other clinical factor is modified due to use of the tool
- [ ] Implementation metrics - metrics related to downloads or interaction with materials guiding people on how to implement or use the tool
- [ ] Fidelity metrics - metrics related to the tool working as expected
- [ ] Satisfaction metrics - metrics related to user (patient, caregiver, physician, etc.) satisfaction with the use of the tool
- [ ] Effectiveness metrics - metrics related to clinical outcomes as related to the use of the tool
- [ ] Cost-effectiveness analysis - metrics related to implementation and maintenance costs as well as expected benefits (e.g., cost savings, lives saved)
- [ ] Evolution metrics - metrics related to changes to the tool or its implementation environment (e.g., hosting) that are necessary for implementation at specific sites
- [ ] Scalability metrics - metrics related to how many different use cases the tool has supported
- [ ] Other: _____

Metrics

In this section, we will ask you questions about more general classifications of metrics that you may have used to evaluate your scientific software/tool.

19. What types of **metrics** have you used to evaluate **user engagement** with your scientific software/tool?

Check all that apply.

- [ ] None
- [ ] Citation Metrics (publication and/or DOI for software directly)
- [ ] Website interaction metrics (number of unique visitors, clicks etc.)
- [ ] Software Downloads (clones, forks, etc.)
- [ ] Documentation engagement metrics (website analytics, video views etc.)
- [ ] Communication engagement metrics (# of emails, survey results, tweets etc.)
- [ ] Development metrics (outside contributions, stars, forks, issues, pull requests)
- [ ] Internal Metrics (new users, registered users, job submissions, error reports etc.)
- [ ] Other: _____

20. How much evaluation of the usage and or impact of the scientific software tool have you done so far?

Mark only one oval.

- ◯ I do not think such evaluations are useful and thus have not performed any
- ◯ I have not attempted any evaluations yet but hope to
- ◯ I monitor basic usage statistics (for example simple download metrics)
- ◯ I regularly perform evaluations involving multiple types of metrics (for example communication metrics, usage metrics, and more)
- ◯ Other: _____

21. Have your evaluations of the user engagement or the impact of your scientific software/tool **influenced** your work?

Check all that apply.

- [ ] None - haven't done enough evaluation yet
- [ ] None - the evaluations haven't been informative enough
- [ ] Informed training/documentation materials
- [ ] Informed outreach strategies to obtain new users
- [ ] Informed performance optimization
- [ ] Informed new development ideas
- [ ] Helped justify funding
- [ ] Other: _____

22. Which of the following metrics have you used to evaluate the **scientific impact** of your scientific software/tool?

Check all that apply.

- [ ] Impact factor or number of citations for papers citing your software
- [ ] Ranking of the use of your tool compared to other similar tools
- [ ] Diversity of usage - different types of journals citing your software, different applications etc.
- [ ] Efficiency/depth metrics - do the papers citing your software require fewer tools or are they able to evaluate a biological phenomenon more extensively than papers using previously available alternatives
- [ ] Discovery - has your software led to new discoveries or terminologies that you can track in manuscripts
- [ ] Other: _____

Metrics that were useful

23. What metrics were especially useful for your evaluations?

_____

24. Please elaborate if you would like.

Motivations

In this next section, we will ask you a few questions about what has motivated your or might motivate you to evaluate user engagement.

25. What aspects of **performance optimization** have you or would you want to learn about?

Check all that apply.

- [ ] unexpected usage patterns or poor adherence to best practices
- [ ] inefficiencies in tool workflows or structures
- [ ] inadequate documentation
- [ ] mismatches between defaults and actual use
- [ ] common errors
- [ ] data volume use
- [ ] None
- [ ] Other:

26. What aspects of **usage optimization** have you or would you want to learn about?

Check all that apply.

- [ ] who users are , where they are, and what they are doing
- [ ] user-base diversity
- [ ] identify sources of other possible users
- [ ] when/where to temper or strengthen user expectations
- [ ] what outreach approaches work best to boost users
- [ ] None
- [ ] Other: _____

27. What aspects of **usability optimization** have you or would you want to learn about?

Check all that apply.

- [ ] what features are often used and by what users
- [ ] what features are not being used
- [ ] if and how users are struggling
- [ ] None
- [ ] Other: _____

28. What **development** aspects have or would be motivational?

Check all that apply.

- [ ] to better understand what data is being used
- [ ] to discover opportunities for new features or data needed
- [ ] to identify more appropriate resource allocation
- [ ] None
- [ ] Other: _____

29. Which of the following have or would be motivational for evaluating your scientific software/tool?

Check all that apply.

- [ ] to support funding requests
- [ ] to support resource requests
- [ ] to promote continued usage by users
- [ ] to promote usage by more diverse users
- [ ] to promote usage of new tools
- [ ] to encourage community contributions
- [ ] None
- [ ] Other: _____

Challenges

We would like to end with some questions about the challenges that you have faced when attempting to evaluate user engagement.

30. What **major barriers** are hindering your ability to evaluate the engagement of your tool(s)?

Check all that apply.

- [ ] None
- [ ] Privacy concerns
- [ ] Security concerns
- [ ] Legal concerns (besides privacy or security)
- [ ] Ethical concerns (besides privacy or security)
- [ ] Technical Issues
- [ ] Not sure what methods to use
- [ ] Limited time to do such analyses
- [ ] Limited funding or other resources
- [ ] Other: _____

31. What **citation challenges** have you encountered?

Check all that apply.

- [ ] None
- [ ] Not applicable
- [ ] Tool is super common so people don't bother to cite
- [ ] Tool only used in discovery phase of research so people forget to cite
- [ ] Hard to track citations of tools that use your tool or are based on your tool
- [ ] Sometimes people cite in location that is difficult to track - abstract or acknowledgments
- [ ] Tool is acknowledged in papers but without formal citation
- [ ] The tool requires hospital/institute support to implement - thus citations aren't a very good estimate of usage
- [ ] Other: _____

32. What **metric distortion challenges** have you encountered?

Check all that apply.

- [ ] None
- [ ] Not applicable
- [ ] Google services or other tracking system banned by some institutions
- [ ] Automations are inflating usage metrics
- [ ] Issues with using software or resources for tracking (ie. github stats)
- [ ] Challenges tracking usage in a cloud environments
- [ ] Distinguishing single users running software many times vs. many users running few times
- [ ] Challenges for large complex projects (software ecosystems)
- [ ] General usage as an (imperfect) proxy for software mature usage
- [ ] Other: _____

33. Is there anything you would like to measure but have been unable to capture?

_____

34. Please elaborate more about any barriers you are experiencing

35. Have you been able to assess your tool's **fairness (Not to be confused with FAIRness** as defined as Findable, Accessible, Interoperable, and Reusable)**.** Here we define **software fairness** in terms of the design being mindful of inclusivity and bias. See this link for more information.

Check all that apply.

- [ ] Not sure how to do such assessments
- [ ] Not sure what this is
- [ ] Attempted, but encountered challenges
- [ ] Yes

Successful evaluation of software fairness

36. Please elaborate about how you successfully evaluated your tool's fairness.

Challenges with software fairness evaluation

37. Please elaborate if you can about any challenges you encountered in evaluating fairness.