Guidelines for Releasing Research Software

RELEASE DATE: DECEMBER 2014

REVISION DATE: April 2015

Filed under: New, Training

Introduction

To encourage open software development by the NCI intramural research community, this document offers guidance for developing and releasing research software source code in a free "open access" or "open source" format. Releasing the software source code in an "open" manner means that you permit users to use the code, modify it, and/or redistribute it. Developers have some flexibility in determining how and when they want to make available software they have developed as part of their research activities.

The answers to the questions in this document reflect NIH policies and guidance for NIH intramural researchers, which promote the broad dissemination of research resources, including research-related software. (See http://grants2.nih.gov/grants/guide/notice-files/not96-184.html; http://sourcebook.od.nih.gov/ethic-conduct/Conduct%20Research%206-11-07.pdf). A guiding principle in this document is the recognition that although under 17 USC §102 copyright protection subsists once a work is fixed in any tangible medium of expression, no copyright protection is available under U.S. law for works of the United States Government, i.e., works authored by U.S. Government employees as part of their official duties (see 17 USC §105). Therefore, NCI employees who develop and release software should do so under an agreement that does not reference U.S. copyright for NCI employees. As discussed below, the model transfer agreement for research software developed by NCI researchers takes this consideration into account.

The following guidelines are relevant to software tools at any level of sophistication, from reusable, "polished" source code to "one-off" code intended for a single, specific project. Reusable source code is designed for multiple uses; reusability encourages distribution and collaboration and is enhanced by application of a user-friendly coding format. One-off code is typically written as an expeditious research method and is intended for single use. One-off code is typically comprised of quick commands and may lack flexibility for alternate uses. If one-off code is required to replicate published findings, then it is important to make the code available, e.g., by deposition in a publicly accessible software repository.

General Questions

1. **Could I be required to make software code I have developed available to researchers upon request?**

Generally, no. However, because NCI is a federal agency, research software code developed by NCI employees in the course of their employment generally should be made available upon request to the research community after its utility has been demonstrated and consistent with any directions from your supervisor. If the code is part of an active research project that has not yet been published, it may not have to be released.

2. **Am I required to actively disseminate code I have developed to the public?**

No. You are not required to actively disseminate the research software code you have developed to the general public. You are however encouraged to make software source code available to the research community as needed in order to better engage in scientific partnership and to accelerate research.

3. **What are the advantages of releasing software source code as "open source" software?**

Releasing software source code contributes to the advancement of science and fosters a spirit of collaboration among researchers across institutions, both academic and commercial. For instance, researchers in the open source software community are often able to identify and resolve software bugs or other issues that could potentially lead to inaccurate findings. Additionally, open source software research tools can be maintained and improved by public contributors. This extends the useful lifetime of the software while also allowing the original developers to dedicate their efforts to other research endeavors.

4. **How do I determine whether NCI has the right to release software source code developed with a contractor?**

If you, as a federal employee, have worked in collaboration with a contractor to produce software code, review the contract and consult with your supervisor and/or the contracting officer to determine the respective rights in the software. If the software has been jointly authored by you and the contractor under a contract that includes Federal Acquisition Regulation (FAR) clause 52.227-14, AND the contract specifically includes software as a covered work, then the U.S. Government retains unlimited rights, i.e., the rights to "use, disclose, reproduce, prepare derivative works, *distribute copies to the public*, and perform publicly and display publicly, in any manner and for any purpose, and to have or permit others to do so" (FAR 52.227-14). Additionally this clause requires the contractor to agree not to assert copyright without the prior written authorization of the Contracting Officer. HOWEVER, if the contractor is authorized to assert copyright in the software, the U.S. Government normally does not have the right to distribute the software. In that case, check with your contracting officer to determine whether it is necessary for the contractor to transfer its copyright in the software to the U.S. Government or whether the U.S. Government is authorized under the contract to distribute the software even if the contractor has been permitted to assert copyright. Although

there is no U.S. copyright in works authored by employees of the U.S. government as part of their official duties, this absence does not limit your ability to distribute software source code that you have developed in the course of your NCI employment.

I am working at NCI as a fellow - what guidelines should I follow for releasing software I developed as open source?

There are many individuals working at NIH who are not employees of the United States Government, such as fellows, consultants, visitors, special volunteers and IPA employees to name a few. Copyright vests in their original works of authorship as soon as the work is fixed in a tangible medium of expression, i.e., it is sufficiently permanent or stable to permit it to be perceived. Sometimes it may be difficult to determine whether an individual is a federal employee, so as a matter of best practice all authors, including federal employees, should be required to assign their copyright rights to the Government before preparation of the work begins. Obviously if the work is a work of the United States Government, not subject to copyright protection, the assignment is a nullity but this is required as a matter of simplicity to avoid having to make a determination of employment, which at times may be difficult. See **APPENDIX A** for a model assignment agreement.

Pre-Release Guidance

5. **How do I release software source code to the public?**

You have several platform options for releasing your software code to the public. Platform options are listed below.

- Release software code via Social Coding sites
    - (e.g., GitHub, Bitbucket, SourceForge, NCIP Hub)
- Release software code to sites supporting the programming language used
    - (e.g., CRAN hosts R, PyPi hosts Python, CPAN hosts Perl)
- Host software code on NCI web server

6. **I would like to use social media tools that encourage collaboration with the public. What is the proper etiquette to follow?**

There are a variety of ways that NCI employees can use NCI or other social media accounts to communicate publicly about research software they have created. Core social media tools include, but are not limited to, the following:

- Social Networks
- Media-sharing Networks
- Blogs
- Social Coding Sites
- Wikis
- Forums

As employees of the federal government, it is important to adhere to the NCI/NIH/HHS policies and guidelines relevant to the use of social media by public servants. These include:

- HHS Guidelines for New and Social Media
- NCI Twitter Guidelines
- Code of Ethics for Government Service
- NIH Principles of Ethical Conduct for Government.

7. **Must I obtain certain NCI permission before releasing software code to the public?**

Unless you are authorized to distribute NCI research tools, you should obtain permission from your immediate supervisor. There are currently no other NCI or NIH permissions necessary to release software code that you have developed to the public at this time. However, the permissions necessary to release software code publicly are subject to change at any time. Furthermore, if the code was developed in collaboration with a contractor or other partner, you will need to consider the terms relating to distribution under the applicable agreement. See the response to **Question #4 above (*How do I determine whether NCI has the right to release software source code developed with a contractor?*).**

8. **Is there a specific time frame by which the code must be released to the public?**

Because you do not have a requirement to release software code you have developed to the public, there is no specific timeframe by which you must release the code. However, in keeping with NIH policies and guidance for conducting intramural research at NIH you should be prepared to release research software which has been discussed in publications or scientific meetings, subject to the need to protect the confidentiality of clinical data or to comply with agreements that preclude dissemination, upon request from members of the research community. Also, be aware that if you refer to the software in a scientific publication, the journal will most likely require that you deposit the code in a publicly accessible archive of some sort.

9. **How thoroughly should I document the code?**

You should document software code that you have developed in enough detail to allow reasonably skilled programmers to update, extend, or replicate your analysis. You should also document the design and purpose of the code and include comments throughout. Taking these measures ensures that you are adhering to best practices and may limit the number of troubleshooting questions sent to you and NCI.

10. **Who is the licensor for software code that I have developed?**

Code developed as part of your official duties as a federal employee is owned by the U.S. Government. As such, the Government is the licensor and the code may be distributed by NCI but not by you in your personal capacity.

11. **Should I distribute code I have developed under a license agreement? If so, why?**

It is important for you to attach a license or other agreement to software source code that you have developed to facilitate the use of the code as intended and to disclaim any liability for any problems associated with the software after the source code has been released to the public. Once software is made available on an "open source" basis, anyone can use, modify, and redistribute the original source code as specified in the accompanying license agreement. The license can be included in the software code as a separate file with a header designating it as the license.

In addition to attaching a license or other agreement to software source code that you develop, it is also important to adhere to the license terms on any external source code that you incorporate into the code, including retaining the appropriate attribution. In this situation, it may be appropriate to include an additional provision in the license for code you have developed to pass through the relevant terms for the externally created software. Please consult with your designated NCI Technology Transfer Center representative if you need further assistance.

12. **How do I choose a license under which to release software code that I have developed at the NCI?**

The open source software community has gravitated toward the use of several standard licenses that have been approved by the Open Source Initiative (OSI). Using one of these licenses makes using and contributing to research software easier because these licenses are well known and understood. Since NCI encourages open source development, NCI CBIIT has adopted the use of the OSI-approved BSD-3 Clause license for all ***third party*** contributions to community-driven software development projects. http://ncip.nci.nih.gov/blog/ncips-open-development-strategy-enables-community-driven-cancer-informatics-software-development/. The BSD-3 Clause license is "permissive" in that it only requires that the original developer(s) and contributor(s) be credited in any distribution or derivative of the software or source code but does not mandate that derivative source code be distributed on an open source basis.

Since this license is based on the ability of the developer(s) to assert U.S. copyright in the software, it is not appropriate for works developed solely by federal employees since software they create in the course of their duties is not protected by U.S. copyright law. Instead, the NCI Technology Transfer Center (TTC) has prepared a model agreement for research software developed by NCI staff (federal employees) that is compatible with the BSD-3 Clause license and also takes into account other concerns that are important to NCI and NIH. See **APPENDIX B** for a copy of this agreement. It is recommended that you use this model agreement for research software you develop as part of your work at NCI. If you believe that you need to modify this model agreement or create a custom agreement, please consult with your designated representative in the NCI Technology Transfer Center. If you have developed the software in collaboration with a contract partner, please consult with your supervisor and/or the contracting officer for guidance as to how the distribution or license agreement should be prepared.

13. **What are my options for restricting access to research software I have largely contributed to or developed?**

Justifications for restricting access to research software that you have developed in whole or in part include:

- The software code is part of ongoing and unpublished research.

- Patent protection for the software is under consideration.

- The software code contains or may be used to obtain personally identifiable information.

If you would like to share such software with another party (e.g., a colleague at NIH or an outside institution) under these circumstances, you may request that your technology transfer office prepare a Software Transfer Agreement (STA). A STA is a contract that can be used to

5

restrict the other party from redistributing the software source code or to limit access to object code with a prohibition on reverse engineering to derive source code. If the other party violates the STA, it may be in breach of contract. Keep in mind that if you have modified software developed by someone else, you would need to pass through any distribution terms associated with the original code. In addition, if you have co-developed the code (either the original or contributed code) with a contractor, the terms of the contract could also affect the terms of the STA. Consult with your designated representative in the NCI Technology Transfer Center for further details and to discuss other potential options for protecting works you have largely contributed to or developed.

14. **How should I acknowledge NIH as the source funding the research software that I have developed?**

Provide the same funding source information as you would in a publication. Be sure to include this information in the software documentation, the license agreement, and the website or the disk/drive from which people download the software. Make sure to identify all contributors to the research software. If the software was developed in collaboration with NCI contract employees, you may need to include a copyright notice alongside the funding source information.   Consult with your supervisor, the contracting officer or your designated NCI TTC representative to address this question.

If you would like include further acknowledgement, include the following comments in your source code:

> \# [Name of software];
> \#Author: [Names of software developers and contributors];
> \# This code was developed through support of the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research [if applicable]
> [Begin Code…]

15. **Can the software user interface contain NCI or NIH logos?**

The Logo/Branding Contact in the NCI Office of Communications and Education (OCE) manages request for use of the NCI, NIH and HHS logos. For more information, please see "NCI Logo & Other Identifying Elements: What You Need to Know". Please consult with the Logo/Branding Contact in the NCI OCE if you have additional questions.

16. **Does NCI or NIH have any requirements related to testing research software I have developed?**

Neither NCI nor NIH has any standard requirements related to software testing. Therefore, you should include a disclaimer on the download page and in the software documentation. Here is the disclaimer used in the model distribution agreement for software developed by NCI TTC for NCI employees:

> THE SOFTWARE IS PROVIDED "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT ARE DISCLAIMED. IN NO EVENT SHALL THE NATIONAL CANCER INSTITUTE (THE PROVIDER), THE NATIONAL INSTITUTES OF HEALTH, THE

U.S. GOVERNMENT OR THE INDIVIDUAL DEVELOPERS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

There are no standard QA/QC steps required before software can be released. Use your professional judgment about the quality of work and have colleagues and collaborators review your work.

17. **How do I check software I have developed for security vulnerabilities?**

If the software you have developed operates as a web application, NCI can scan it for vulnerabilities using the Institute's security scanning tool. In addition, the SANS institute has published a list of the top 25 software errors. These lists could help guide you in review of the software source code. See http://www.sans.org/top25-software-errors/.

18. **Can research software I have developed be used in medical practice or clinical settings? What additional steps should I take before release?**

In most instances, research software you develop will not be used for clinical or medical purposes, which is why the model software agreement in **APPENDIX B** states that "THE SOFTWARE SHALL NOT BE USED IN THE TREATMENT OR DIAGNOSIS OF HUMAN SUBJECTS." The model agreement in **APPENDIX B** also includes a detailed disclaimer. If you know that the software is intended for use in a clinical or medical setting, please consult with your supervisor and/or your program's senior leadership so that the software may undergo appropriate scientific and regulatory reviews.

Post-Release Guidance

19. **Am I expected to provide technical support in perpetuity to software users?**

Ultimately it is up to the original developer(s) to decide whether to provide technical support (e.g., fix bugs or answer questions) and to what extent. Depending on the platform used to release the code, community contributions can be supported and the decision should be primarily driven by research needs and use. The developer(s) should explicitly state in a "README" file whether the source code is unsupported or supported and, if supported, for how long and how to contact the developer(s).

20. **How can I retain control over open source software?**

In short, you cannot control what other people do with the software once you have distributed it under an open source software license or other agreement. As defined by the Open Source Initiative (http://opensource.org/osd), open source software means that the original source code

is made freely available and may be redistributed and modified. The model software agreement in **APPENDIX B** does not require that subsequent users of the software make their modifications openly available. While there are other license models that do mandate open source distribution of subsequent code contributions, for example, the GPL (GNU) license, it is predicated on the code being copyrighted. However since copyright protection is not available for software authored by U.S. Government employees, the use of GPL (GNU) license by NCI employees is generally discouraged.

If this is an important concern, your best option is to continue actively developing and updating the software, and to focus on remaining as the leading expert in the software's scientific area. The open source community typically respects such dedicated stewardship of a project. In addition, you can accept offers from other developers to help maintain or extend the software. You may find productive collaborators this way.

21. **Can I restrict who can access and use the software (u.g., for data sensitivity reasons)?**

No, you are not able to put restrictions on who can access and use the software if you are releasing it under a free "open source" or "open access" format. Please see the above response to **Question #13** (***What are my options for protecting research software I have largely contributed to or developed?***). If the software contains sensitive information, it would not be appropriate to release it under an open source agreement. In that case, please consult with your designated representative in the NCI Technology Transfer Center.

22. **How can I reference the software in a publication so that readers may access it?**

NCI/NIH does not have a standard convention for referencing specialized software. The APA (American Psychological Association) citation guide recommends the following format:

Author. (Date). Title of Software or Computer Program (Version no.) [Any identifier]. City, state: Publisher. Retrieved Month day, year. Available from URL.

23. **How can I require acknowledgment from people publishing research based on the software or software derived from it?**

Although NCI (or NIH) by itself cannot enforce a license term that *requires* acknowledgement from people publishing research based on the software or its derivatives, you may *request* acknowledgement in a README file associated with the software code. For example, you could include the following statement: "Please cite the author in any work or product based on this software." This point is also specifically addressed in paragraph 3 of the model software agreement in **APPENDIX B.** In addition, you should explicitly describe how to properly cite the source if this is of concern.

24. **After the software has been released, how can I notify users of updates? And, what responsibility do I have for maintaining and updating the software?**

You do not have an obligation to maintain and update the software. The amount of time and effort you commit to maintaining and updating the software is driven by the research that led to the development of the software. Social coding environments like GitHub offer the ability for interested community members to "follow" the project and receive automatic updates about

project activity. The open source community respects dedicated stewardship of a project. You may accept non-monetary offers from other developers to help maintain or extend your software. Often, if there is a dedicated following, supporting developers may take over development and you may step down. Be mindful that if no one takes over development, your software may become outdated.

You may also maintain a dedicated webpage for the software or one that catalogs software and notifies users of availability and updates. For example, see http://dceg.cancer.gov/tools and http://cbiit.nci.nih.gov/ncip/biomedical-informatics-resources/applications.

ADDITIONAL RESOURCES

- **APPENDIX A**: Copyright Assignment
- **APPENDIX B**: Open Source Software Distribution Agreement (for Software Developed by NCI Staff (Federal Employees))

Primary NCI Technology Transfer Contact (for NCI CBIIT):
Ms. Wendy Patterson, Esq.
Technology Transfer Advisor
NCI Technology Transfer Center, NCI Office of Management
pattersw@mail.nih.gov

APPENDIX A: <u>Copyright Assignment</u>

I, _____ (*name*) hereby assign the entire right title and interest in and to the copyright in any original work of authorship made during my participation in _____ (*describe project*) to the United States Government.

BY: _____   DATE: _____
     (SIGNATURE)

NAME: _____
TITLE: _____
ORGANIZATION: _____
DIV/BRANCH/LAB: _____

APPENDIX B: Open Source Software Distribution Agreement (for Software Developed by NCI Staff (Federal Employees))

XXXX, v._____   <insert name and version number of software> (Release date: ____ <insert date>)

- **DEFINITIONS:** AUTHOR(S) XXXX *<insert name, employer, affiliation as appropriate>*

- PROVIDER: the National Cancer Institute (NCI), a participating institute of the National Institutes of Health (NIH), and an agency of the United States Government.

- SOFTWARE: the human readable source code form, the machine readable, binary, object code form, and the related documentation for the modules of the XXXXX*<insert name and version number of software>* software package, which is a _____ *<insert brief description of software's purpose>*.

- RECIPIENT: the party that downloads the software.

By downloading or otherwise receiving the SOFTWARE, RECIPIENT may use and/or redistribute the SOFTWARE, with or without modification, subject to RECIPIENT's agreement to the following terms:

1. THE SOFTWARE SHALL NOT BE USED IN THE TREATMENT OR DIAGNOSIS OF HUMAN SUBJECTS. RECIPIENT is responsible for compliance with all laws and regulations applicable to the use of the SOFTWARE.

2. The SOFTWARE that is distributed pursuant to this Agreement has been created by United States Government employees. In accordance with Title 17 of the United States Code, section 105, the SOFTWARE is not subject to copyright protection in the United States. Other than copyright, all rights, title and interest in the SOFTWARE shall remain with the PROVIDER.

3. RECIPIENT agrees to acknowledge PROVIDER's contribution and the name of the author of the SOFTWARE in all written publications containing any data or information regarding or resulting from use of the SOFTWARE.

4. THE SOFTWARE IS PROVIDED "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT ARE DISCLAIMED. IN NO EVENT SHALL THE PROVIDER OR THE INDIVIDUAL DEVELOPERS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

5. RECIPIENT agrees not to use any trademarks, service marks, trade names, logos or product names of NCI or NIH to endorse or promote products derived from the SOFTWARE without specific, prior and written permission.

6. For sake of clarity, and not by way of limitation, RECIPIENT may add its own copyright statement to its modifications or derivative works of the SOFTWARE and may provide additional or different license terms and conditions in its sublicenses of modifications or derivative works of the SOFTWARE provided that RECIPIENT's use, reproduction, and distribution of the SOFTWARE otherwise complies with the conditions stated in this Agreement. Whenever Recipient distributes or redistributes the SOFTWARE, a copy of this Agreement must be included with each copy of the SOFTWARE.