CENTRE FOR GENOMIC REGULATION

SCIENTIFIC SOFTWARE POLICY

October 2014

INTRODUCTION

The mission of CRG is the gathering and dissemination of knowledge for the benefit of society, public health and economic prosperity. CRG as a public institution has an obligation/responsibility to make research results available for the benefit of the community. Developing software (including research tools) and putting it to use in the service of the public is a thoroughly valid mode of accomplishing this objective.

This Scientific Software Policy adheres to the recently published PLoS data and software policies and Nar policy on web servers and describes the corporate rules affecting the commitments and duties of the scientists, the group leaders, the core facilities and the departments of the Management area during the Application Lifecycle Management (hereafter ALM).

By default, any piece of scientific software developed in the CRG will be delivered as Open Source software and distributed under Open Source Licenses (OSL) or any other suitable standard license (see http://opensource.org/licenses/alphabetical), like Academic Free Licenses (AFL). Dependencies on commercial software are allowed and are to be properly documented, although complete open source solutions are preferred.

In some cases, it can be decided that one or some pieces of software will be delivered as Proprietary Software Licenses or as Commercial Software Licenses.

It will be allowed that the same piece of software is distributed under both AFL and Proprietary /Commercial licensing model at the same time.

This document is shared by the Information Technologies Department (hereafter IT), the Technology and Business Development Office (hereafter TBDO), the Scientific Information Technologies Unit (hereafter SIT) and the Bioinformatics Core Facility (hereafter BioCore).

APPLICATION LIFE CYCLE POLICIES AND PROCEDURES IN THE CRG

This Chapter describes the role of every department, unit and group involved in the application lifecycle management (ALM) at the CRG.

Out of simplicity, this document will refer to the classical definition of ALM () in a waterfall methodology.

The following figure enumerates the ALM phases/modules:

Portfolio Management, Change Management, Project Management and Process and Best Practices define the different high level processes and activities related to the software strategy, providing a framework for decision making on creation/maintenance/deprecation of an application, versions scheduling, dissemination and marketing plans and portfolio management, etc.

On the other hand, Requirements, Design and Development, Build Workflow and Testing/Validation phases are the low-level parts of the so-called Application Development Lifecycle Management (hereafter ADLM), it is, the process of developing and maintaining a specific application or piece of software.

(a) Scientific Groups

The **Scientific Group** that will develop the application will be fully responsible for its ADLM. Specifically, the Group Leader or his/her representative for software development will also be responsible for:

1. Communicating to the BioCore the list of validated software his group developed in the CRG. Only correctly tested software should be considered here as they will be included into the CRG software catalogue. By default, all software are distributed under an open source license. However, if the group leader needs advices or decide to apply a commercial license, he must contact the TBDO. In that case, the scientists and the TBDO will decide whether the software should be open source, it should be license at no cost for academic research institutes (only for research purposes) or if it is better to obtain a particular IP protection or to share it under a commercial license.

2. Keeping the complete source code and all the required libraries, release management scripts and in general all the related documentation in a public repository, so that it may be kept safe and re-generated if needed. Should the software be of any commercial interest the software will be place at a CRG repository maintained by the IT dept. *Note: Public repositories are recommended as specified in* *http://www.plosone.org/static/policies%20-%20sharing#sharing*

3. Only for academic and commercial software, providing to the IT dept. the required information for the public software download web page, as defined by the IT dept. through the new CRG software definition specs form.

4. Providing to the BioCore the minimum required information for the public web page (see Annex 3: Sample for application webpages)

5. Whenever it is possible, providing Galaxy wrappers for the BioCore to integrate them in the corporate Galaxy instances.

6. Fulfilling steps 1 to 6 for any new version to be delivered, so that the ALM is kept up to date at any time.

7. Once a year, providing the updated information on the software developed by his/her group, including at least:
   i. Latest versions of the software (to the BioCore);
   ii. Validity / expiration date of the previous versions (to the BioCore);
   iii. Number of downloads per version in case of software commercialization (to TBDO).

8. Developing and maintaining every specific web page for any software either totally or partially developed in the CRG. These web pages will always include the CRG logo and will comply with the rest of corporate requirements set by the Communications dept.

9. Assuring future accuracy of the software developed by his/her group with regards to the PLoS data and software policies and with regards to the Nar policy on web servers.

(b) The BioCore

The **BioCore** will be responsible for the following duties:

1. Together with the ITC department it will maintain a template form for the CRG software definition specs form where to collect at once all the required info about any software from the group leader.

2. It will provide on demand support for Web sites and eventual related databases at internal price rate.

3. It will provide on demand support for Galaxy wrapper development at internal price rate.

4. Together with the IT department, it will maintain a public webpage with all the application files (see Annex 3: Sample for application webpages).

5. It will maintain an internal mailing list for scientific software support.

6. It will provide on demand support for code development by assessing in the ALM phases, as well as on the use of mapping databases.

7. It will develop modules and APIs on demand.

8. It will run a yearly update of the public webpage.

9. In case an additional resource is provided, it will maintain and animate a public wiki/forum webpage supporting installation and providing tutorials and best practices for development

(c) The Information Technology (IT) department

The **IT dept.** will be responsible for the following duties:

1. It will provide virtual machines and hosts production servers for web pages and all the services associated with the scientific software, except for High Performance Computing (HPC) and storage (see SIT section).

2. It will set and maintain templates for several types of Virtual Machines according to the software requirements.

3. For Academic and Proprietary/Commercial licensed software, it will administer and maintain the unique common public web page for registration and code downloads.

(d) The Scientific Information Technologies (SIT) Unit

The SIT unit will be responsible for the following duties:

1. It will provide support for High Performance Computing (HPC) and for storage servers where scientific data are housed.

2. It will support and hosts the necessary MySQL database servers in production.

(e) The Technology and Business Development Office

The mission of CRG is the gathering and dissemination of knowledge for the benefit of society, public health and economic prosperity. CRG as a public institution has an obligation/responsibility to make research results available for the benefit of the community. Developing software (including research tools) and putting it to use in the service of the public

is a thoroughly valid mode of accomplishing this objective. Entering commercial arrangements is one way by which the Institute can ensure that the technology is made available and protected at the same time. This transfer of technology to industry increases the researcher's exposure to that industry´s expertise and needs, thus benefiting the researcher, the Institute, and the Industry. Finally, successful arrangements will generate income for the Institute and researcher, and enhance their reputation. Accordingly, CRG encourages the inventive process and can provide advice and assistance in bringing software to the point of public use.

The Software Disclosure Form (SDF) is the tool provided by the TBDO to assist you in the Software disclosure required by the CRG's Intellectual Property Policy. The purpose of disclosing a Software to the TBDO is to determine copyright ownership and to explore licensing opportunities. SDFs are legally important documents, which should be read carefully and filled out completely. All personal information given to TBDO is considered confidential. This information will not be sold, posted, made publicly available or shared with third parties.

Copyright is a form of intellectual property protection granted to authors of original works that are fixed in a tangible medium of expression. The medium need not be directly perceptible so long as it may be communicated with the aid of a machine or device; therefore, both computer software and documentation are copyrightable. A copyright owner has certain exclusive rights, including the right to reproduce, distribute, prepare derivative works, and display the work, and may authorize others to exercise those rights too.

When to submit an SDF

Software should be disclosed to the TBDO before it is distributed outside the CRG, so that the TBDO can determine under which conditions the software can be legally distributed. In addition, disclosure might be required by the contract under which the software was developed, or might be required by a co-owner of the software.

In the sense used here, software has presumed commercial use and value. The following indicates the preferred time to file a disclosure with the TBDO:

- **Consult with TBDO:** TBDO can assist you in determining inventorship and authorship, implementing procedures when more than one institution is involved, and protecting patentable software from disqualifying public disclosure.
- **Disclose to the TBDO first, publish later:** Disclosing your software to the TBDO by no means prescribes publication. On the other hand, if there is patentable subject matter within or associated with the software, premature publication may preclude the availability of patent protection in most countries.
- **Disclose to the TBDO as soon as the software is clearly conceptualized**: It is not necessary to wait until the software is reduced to practice. Where patent applications are involved, filing early may be beneficial, particularly when others are filing related patent applications.

Considerations about Licensing and Open Source Distribution

There are many different types of software licenses. The optimal license depends on a variety of factors including the form in which the code is being released (source or object), the rights that licensee will have in the software, and whether the software is protected by patent as well as copyright.

The TBDO reviews all SDFs as they are received from members of the CRG community and will work with the authors/inventors to determine the optimal licensing program. The TBDO will assist in licensing your software:

- For academic and research purposes only, via an open source or free distribution license, or
- To industry for further development and commercialization. Any royalties derived from any such license are shared with the author(s) and their departments according to CRG Policy.

At the request of the author(s), as evidenced by their signatures on the Form, CRG will distribute software via an open source license, without fee or royalty. The TBDO supports this approach if the authors of the software feel it is an appropriate distribution method for the software in question and such distribution has been approved by the head of the relevant department, laboratory, or center.

In open sourcing your software, you must be sure that this is OK with any sponsors of your research and that there is not an active sponsored research grant that would prevent such distribution. A sponsor may have rights in the software as a result of their research funding and may not want the software's source code to be made open.

Further, if your software contains other "open source" or "free" software, or any software that you have downloaded from or been provided by a third party, you must understand the license terms under which that software has been provided. Certain license terms have specific restrictions. The TBDO can assist in understanding the different terms.

Once third party rights, if any, are resolved, there are numerous approved open source license models from which to choose. "Approved" open source licenses are those approved by the Open Source Initiative, however, any license that releases the source code without a required payment is an open source, or sometimes-called "shared source," license. A complete list of available licenses can be seen at www.opensource.org/licenses. Generally, CRG recommends either the BSD license or the GPL or LGPL licenses:

- BSD License (http://www.opensource.org/licenses/bsd-license.php): In this case, CRG retains copyright ownership, but makes the software openly available to the public without the requirement that redistribution of the software include the source code. This is the least restrictive open source license. If CRG approves the open source distribution, the license template at the referenced website should be included conspicuously within the released software copy and at the site where the software is downloaded.

- GPL or LGPL License (http://www.opensource.org/licenses/gpl-license.php and http://www.opensource.org/licenses/lgpl-license.php): In this case, CRG retains copyright ownership, but makes the software openly available to the public with the requirement that redistribution of the software, in whole or in part (or any modification of or addition to the software) include the source code. The referenced web sites contain instructions on how to release software in accordance with those license terms.

CRG recognizes that other open source licensing mechanisms exist. If you desire a different open source licensing strategy, please confer with a TBDO Manager.

The Software/Copyright Disclosure Form (SDF) is the first step in the process of identifying copyrightable software and determining how to proceed with the resulting Intellectual Property (IP). This determination process includes assessments about ownership, encumbrances and commercialization. Researchers should provide as much information about the Software as possible. The TBDO will use this Information in assessing the appropriate means for protecting the software (copyright, patent) and identifying possible funding, sponsorship and licensing opportunities. Upon receipt, the TTO will review the form and contact you within a few days to further discuss the disclosure.

Copyright does not protect ideas, procedures, methods, systems, processes, concepts, principles, discoveries, devices, formulas or algorithms (as distinguished from the way these things are expressed, described or illustrated). Although not protected by copyright, these discoveries may be the subject of patent protection.

For software, the TBDO does not generally file provisional patent applications. If you think that the software described below may implement a patentable invention, you must ALSO fill out an Invention Disclosure Form (see: http://www.crg.eu/) for the patentable aspects of the software (e.g. an algorithm or method) and submit these forms together. If both forms are used, they should reference each other in the description of the software. Inventors of the process need not be (and likely are not) identical to the authors of the code.

Annex 1: sample for application webpages

For every application developed and delivered, the Group Leader will provide the BioCore with the following information:

- Name
- Category
- Short description
- Link to the specific web page
- Link to the latest versions
- *(supported list of)* Operating system
- Last update
- Laboratory
- Contact name for technical questions
- Galaxy wrapper (if available)
- Link to the user's guide manual. To include at least:
- Method
- Application
- Tutorial
- Dependencies on commercial software

In general, the BioCore will be entitled to request any information required to fulfill the PLoS data and software policies and the Nar policy on web servers.

With the collected information, the BioCore will maintain the application webpages. See the following samples:

http://biocore.crg.cat/wiki/ExampleCRGSoft

http://biocore.crg.cat/wiki/T-coffee