

Ten Simple Rules

for the Open Development of Scientific Software

Andreas Prlic, Jim Procter, Hilmar Lapp

Citation: Prlić A, Procter JB (2012) Ten Simple Rules for the Open Development of Scientific Software. PLoS Comput Biol 8(12): e1002802. doi:10.1371/journal.pcbi.1002802

- Rule 1: Don't Reinvent the Wheel
- Rule 2: Code Well
- Rule 3: Be Your Own User
- Rule 4: Be Transparent
- Rule 5: Be Simple

- Rule 6: Don't Be a Perfectionist
- Rule 7: Nurture and Grow Your Community
- Rule 8: Promote Your Project
- Rule 9: Find Sponsors
- Rule 10: Science Counts

Rule 5: Be Simple

Be Simple - For yourself and everyone else

- Minimise your timesinks
 - Infrastructure
 - Tests are not enough - use a build system
 - PACKAGING
 - Tests == examples for you and everyone else

- Documentation == outreach
 - love your website
 - love your programmer docs
 - issues == documentation

Why not start your next grant proposal on

- <insert ovcs here> ?

Rule 1: Don't Reinvent
the Wheel

All Requests 25

Open Closed Sort: Newest ▾

New pull request

Yours

Find a user...

bow	3
diegozea	2
timwintle	1
JoaoRodrigues	1
msabramo	1
hammer	1
clparallel	1
kevinwuhoo	1
bendmorris	1
saketkc	1
polyatail	1
konstantint	1
DavidCain	1
sefakilic	1
anntzer	1
kdmurray91	1
battmatt	1
davidmam	1
peterjc	1
nmsutton	1
MarkusPiotrowski	1
matsu3shiro	1

-  **Windows line ending issue fixes** #197

Some fixes from Codefest.

 by konstantint 13 hours ago  1 comment
-  **Fixing a bug in SeqIO.PdbIO in parsing pdb-atom sequence** #196

Fixing a problem that SeqIO.PdbIO outputs erroneous amino acid sequence if the ATOM record has mo...

 by matsu3shiro 16 days ago  2 comments
-  **MeltingTemp completely rewritten and extended** #192

More or less completely rewritten and largely extended. 1. Three different Tm calculations: one ...

 by MarkusPiotrowski 20 days ago  1 comment
-  **CDAO: removing dependency on librdf** #193

Users of the Bio.Phylo CDAO conversion modules have had problems installing its dependency, the R...

 by bendmorris 20 days ago
-  **Customizability features added to Phylo.draw** #189

Features added based on <https://redmine.open-bio.org/issues/3336>. Users now have the ability to i...

 by nmsutton a month ago
-  **Samtools Wrapper For BioPython** #180

Added Samtools(<http://samtools.sourceforge.net/>) wrapper. Need to write Unit Tests.

 by saketkc 2 months ago  11 comments
-  **Reorganization of single and triple letter protein codes** #174

This pull request was triggered by the discussion here. Summary of the changes: I've moved the SC...

 by bow 4 months ago  2 comments
-  **KEGG KGML parser and graphical visualisation from Leighton** #173

This pull request is mainly as a discussion point for Leighton's (@widdowquinn) KEGG KGML parser ...

 by peterjc 4 months ago  4 comments
-  **Add the ability to parse CEL version 4 files from Affy** #168

Hey, I noticed that Biopython was missing the ability to parse binary CEL files (version 4), so I...

 by hammer 4 months ago  8 comments
-  **Change to set_structure. Allows writing of individual (S)MCRA's** #164

The changes discussed in this thread: <http://lists.open-bio.org/pipermail/biopython-dev/2013-Febr...>

Share the full story of your research impact.

ImpactStory is your impact profile on the web: we reveal the diverse impacts of your articles, datasets, software, and more.

[Make my impact profile](#)

[or, show me a sample profile](#)



View metrics in seconds

Point us to your slides, code, datasets, and articles. In a few seconds, you'll have a report detailing your impacts: citations.



Embed them anywhere

Drop ImpactStory's embeddable **Javascript widget** into your own online CV or website to show the impacts of your projects.




Open data, open source.


Our data, like our **source code** wide open. As a non-profit, we around supporting open tools nurture Web-native scholarship.





Peter Cock

 Loading...

 add products

 refresh metrics

 download 

software

biopython

(2009) My fork of the official Biopython repository, used for experimental branches etc [GitHub](#)

maf2sam

(2010) Convert MIRA Assembly Format (MAF) to Sequence Alignment/Map (SAM) format [GitHub](#)

picobio

(2011) Miscellaneous Bioinformatics scripts etc mostly in Python [GitHub](#)

tarball2git

(2011) Simple Python script to take a set of versioned tar balls and import them into a git repository [GitHub](#)

screed

(2010) a short read database [GitHub](#)

longsight

(2012) Python code for capturing images from a webcam etc [GitHub](#)

backports.lzma

(2012) Backport of Python 3.3's standard library module lzma for LZMA/XY compressed files [GitHub](#)

split-dist

(2011) Unofficial repository for Thomas Mailund's tool Split-Dist (sdist) [GitHub](#)

ccp

highly recommended

highly cited

Highly recommended by the public

This item has 12 GitHub stars. That's better than 87% of items added to GitHub in 2009, suggesting it's highly recommended by the public. Click to learn more.

recommended

cited

discussed

cited

recommended

cited

recommended



Hilmar Lapp

hlapp

<> PHP, Java

📍 Durham, NC

✉ hlapp@drycafe.net

🕒 Joined on Feb 03, 2009

27

followers

27

starred

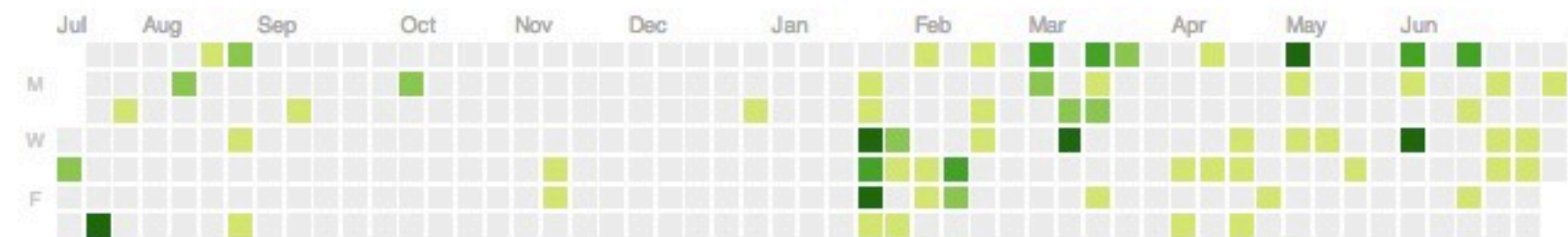
2

following

Organizations



Your Contributions



🔍 Summary of Pull Requests, issues opened and commits. [Learn more.](#)

Less More

293 Total

Jul 18 2012 - Jul 18 2013

Year of Contributions

6 days

January 28 - February 02

Longest Streak

0 days

Rock - Hard Place

Current Streak

Contribution Activity

Period: 1 Month ▾

🔗 7 Commits

[Pushed 2 commits to phenoscape/phenoscape.github.com](#) Jul 15

[Pushed 3 commits to datadryad/dryad-data](#) Jul 10

[Pushed 1 commit to bendmorris/phylocommons](#) Jun 25

[Pushed 1 commit to phylotastic/phylotastic.github.com](#) Jun 23

🔗 3 Pull Requests

merged [Revert NCBI Linkout stats deposition.](#) 8 days ago

datadryad/dryad-repo · 1 commit · 0 ++ 367 --

merged [Consolidate and streamline status descriptions and voting process.](#) 17 days ago

OBF/obf-docs · 1 commit · 46 ++ 57 --

closed [Consolidated and streamline status descriptions and voting process.](#) 20 days ago

OBF/obf-docs · 1 commit · 46 ++ 57 --

🔍 12 Issues Reported

closed [#21 Add link to DeepFin RCN](#) 3 days ago

open [#7 Why not distribute under CC0?](#) 7 days ago



New: CloudBioLinux packages are now listed on biogems.info !

#	biogem	description	by	cite	version	released	stars	issues	source	build	total	90d*)	7d	90d**)
1	bio	Bioinformatics library (...)	BioRuby project		1.4.3.0001	7 weeks		...		build passing	441	dow	2013-07-18	AR and MA simulations (GSoC)
2	biodiversity	Parser of scientific names (...)	Dmitry Mozzherin		3.1.2	3 weeks	9	5		build passing	227		2013-07-17	bio-krona 0.1.0 release
3	bio gem	Biogem is a software generator for Ruby in (...)	Raoul J.P. Bonnal, Pjotr Prins		1.3.4	15 months		...		build error	196		2013-07-16	First Four Validations (Monica's GSoC)
4	bio samtools	Binder of samtools for Ruby, on the top (...)	Ricardo Ramirez-Gonzalez, Dan MacLean, Raoul J.P. Bonnal		0.6.0	7 months	13	4		build passing	152		2013-07-12	bio-raxml 0.1.0 release
5	entrez	Http requests to entrez e-utilities (...)	Jared Ning		0.5.8.1	23 months		...		build unknown	130		2013-07-12	Cubecumber (Will's GSoC)
6	bio ucsc api	The Ruby ucsc api: accessing the ucsc genome (...)	Hiroyuki Mishima, Jan Aerts		0.5.2	7 months		...		build passing	113		2013-07-11	How Plotrb Works Inter (Zuhao's GSoC)
7	intermine	Webservice client library for intermine data-warehouses (...)	Alex Kalderimis		1.04.00	1 week		...		build error	109		2013-07-10	Cuke up with Cucumber (GSoC)
8	bio gff3	Gff3 parser for big data (...)	Pjotr Prins		0.9.1	11 months		...		build passing	104		2013-07-09	Ruby mining: first draft (Alberto's GSoC)
9	sequenceserver	Blast search made easy! (...)	Anurag Priyam, Ben J Woodcroft, Yannick Wurm		0.8.5	11 weeks		...		build passing	90		2013-07-09	Imperfect Forward Secrecy (Coming Cryptocalypse) (Arcieri)
10	bio logger	Log4r wrapper with extra features for roles and (...)	Pjotr Prins		1.0.1	15 months	4	2		build passing	78		2013-07-09	Some Thoughts on Ruby's attr_reader (Zuhao's GSoC)
11	bio maf	Maf parser for BioRuby (...)	Clayton Wheeler		1.0.1	11 months	9	29		build passing	74		2013-07-08	Coded properties for gene semantics (Will's GSoC)
12	bio gadget	Gadgets for bioinformatics (...)	Shintaro Katayama		0.4.8	5 weeks		...		build error	65		2013-07-07	intermine 1.04.00 release
13	bio grid	A biogem to submit jobs on a queue (...)	Francesco Strozzi		0.3.3	8 months		...		build unknown	59		2013-07-07	Problems with predicted gene merges (Monica's GSoC)
14	bio blastxmlparser	Very fast blast xml parser and library for (...)	Pjotr Prins		1.1.1	5 months	9	...		build passing	57		2013-07-06	Databases for predicted (Monica's GSoC)
15	bio faster	A fast parser for fastq files (...)	Francesco Strozzi		0.4.5	13 months		...		build passing	53		2013-07-06	Statsample - Partial Autocorrelation (Ankur)

Rule 10: Science Counts

- Differences between computer sciences and bio-sciences
- Software is primarily a means to advance our research
- Software for the consumption of others does not get rewarded
- We write open source because we like to interact with other scientists

- “Build it and they will come” does not work
- Maintenance of code that is no longer relevant to your own research is a serious time sink
- If done right, you can publish both the science and the software for the same project

- Question:
- What do you need to do to get
 - Published
 - Grants
 - Academic Promotions
 - Tenure ?

- Software Articles
- <http://www.ploscompbiol.org/static/guidelines#software>

Call for Papers

First Workshop on Sustainable Software for Science: Practice and Experiences (WSSSPE)

<http://wssspe.researchcomputing.org.uk/>

(in conjunction with SC13)

Sunday, November 17, 2013, Denver, CO

Progress in scientific research is dependent on the quality and accessibility of software at all levels and it is now critical to address many new challenges related to the development, deployment, and maintenance of reusable software. In addition, it is essential that scientists, researchers, and students are able to learn and adopt a new set of software-related skills and methodologies. Established researchers are already acquiring some of these skills, and in particular a specialized class of software developers is emerging in academic environments who are an integral and embedded part of successful research teams. This workshop will provide a forum for discussion of the challenges, including both positions and experiences. The short papers and discussion will be archived as a basis for continued discussion, and we intend the workshop to feed into the collaborative writing of one or more journal publications.

In practice, scientific software activities are part of an ecosystem where key roles are held by developers, users, and funders. All three groups supply resources to the ecosystem, as well as requirements that bound it. Roughly following the example of NSF's Vision and Strategy for Software (http://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf12113), the ecosystem may be viewed as having challenges related to: