Collaborative Software Development
Using R-Forge

Stefan Theußl    Achim Zeileis    Kurt Hornik

Department of Statistics and Mathematics
Wirtschaftsuniversität Wien

August 13, 2008
Why Open Source?

- Source code is by definition available to everyone
- Reuse of existing code
- Rapid creation of solutions within an open environment
  “Release early, release often” paradigm (Linus Torvalds)
- Peer review of open source software (OSS)

A key to the success of open source projects is **collaboration**
Collaboration in the R Community

- For a decade, the R Development Core Team is using development tools like Subversion (SVN) for managing source code.
- A central repository is hosted at ETH Zürich for managing the development of the base R system.
- Many package developers use similar infrastructure to manage their source code.
- Around 46.8% of the 1500 packages available on CRAN are authored by more than two developers.

<table>
<thead>
<tr>
<th>Number of authors per package in %</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>≥ 8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>23.9</td>
<td>11.6</td>
<td>6.3</td>
<td>2.4</td>
<td>1.3</td>
<td>0.4</td>
<td>0.8</td>
</tr>
</tbody>
</table>
Source Code Management

Why do developers use source code management (SCM) tools?

- Efficient collaboration and sharing knowledge
- Easy communication through various channels
- Shared storage for source code
- Version control
- Larger software projects can be managed more efficiently
- Users may participate and give feedback
Introduction to R-Forge

What is R-Forge?

▶ A central platform for the development of R packages, R related software and further projects
▶ R-Forge can be found at http://R-Forge.R-project.org
▶ R-Forge offers several tools to help package developeRs to collaborate

Since starting the platform in early 2007 more and more interested useRs registered projects on R-Forge. Now after a year being in a development and testing stage nearly 200 projects and around 500 useRs are registered on R-Forge.
R-Forge is based on GForge

What is GForge? Why is R-Forge based on GForge?

- GForge, a fork of the 2.61 SourceForge code, is an open-source project (http://gforge.org)
- GForge employs a php-postgresql framework to offer various tools for collaboration and source code management
- One of the most important reasons using GForge: It allows for the development and usage of plugins
Core Features of R-Forge

- Source code management with SVN
- A CRAN-style repository for hosting development releases of R packages
- Daily built packages are available for Linux, Mac OS X and Windows
- Packages can be downloaded from the website and/or installed in R via `install.packages("foo", repos = "http://R-Forge.R-project.org")`
Additional Features of R-Forge

- Mailing lists
- Discussion forums
- News announcements to be posted on homepage
- Project websites (http://foo.R-Forge.R-project.org)
- Project categorization
- Full repository backups
- And many more
Figure: Homepage of R-Forge
Everything on R-Forge is organized in so-called projects.

Every Project has an SVN repository containing two pre-defined directories:

- `pkg` ... contains one or more R packages
- `www` ... optionally contains a Project specific website

Each member of a project is assigned a role (e.g., “Administrator” or “Developer”) with certain rights like write access to the repository, releasing packages to CRAN, ...
Figure: Project Summary Page
Release and Quality Management

- Early versions of software projects are typically prototypes and therefore are not completely bug free.
- Therefore, R-Forge offers a quality management system similar to that of CRAN.
- In the spirit of OSS—“given enough eyeballs, all bugs are shallow” (Eric S. Raymond)—R-Forge additionally provides a bug tracking system for peer code review.
- Eventually, packages passing `R CMD check` on R-Forge can be directly released to CRAN.
Contributed R Packages

Below is a list of all packages provided by Rmetrics - Computational Finance.

Important note for package binaries: R-Forge provides these binaries only for versions of R that were released since the package was first submitted to R-Forge, but not for older versions. In order to use more recent packages from R-Forge you may need to switch to a newer version of R or, alternatively, try to build the package with an older version of R.

For more information on Rmetrics - Computational Finance, please visit their website at http://www.rmetrics.org.

Figure: R Packages Tab
Outlook and Future Work

- Introducing a file similar to DESCRIPTION for controlling the behaviour of the build/check process, etc.
- CRAN-style check summaries
- Automated check result delivery to mailing lists
- Task management, shared TODO lists
- Project wikis
- Sustainable improvement of the R-Forge manual
Outlook and Future Work

Other features demanded by users:

- Inclusion of BioConductor and Omegahat repositories in build/check process
- Controlling of build/check process (e.g., via options passed to \texttt{R CMD check}, . . .)
- Altering changelog messages in SVN repository
- Project renaming

Open bugs:

- \texttt{Tcl/Tk} is not (yet) supported on our Mac
- Build/check process stability and performance
Contact

▶ Regarding R-Forge
email: R-Forge@R-project.org
Support tracker:
https://R-Forge.R-project.org/tracker/?group_id=34
▶ For everything else please contact me directly
Stefan Theußl
Department of Statistics and Mathematics
Wirtschaftsuniversität Wien
Augasse 2–6, A-1090 Wien
Austria
e-mail: Stefan.Theussl@wu-wien.ac.at
URL: http://statmath.wu-wien.ac.at/~theussl