



# 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

Number of authors per package in %						
2	3	4	5	6	7	≥ 8
23.9	11.6	6.3	2.4	1.3	0.4	0.8



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





## 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 OSX 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





### What are R and R-Forge?

R is "GNU S", a freely available language and environment for statistical computing and graphics which provides a wide variety of statistical and graphical techniques: linear and nonlinear modelling, statistical tests, time series analysis, classification, clustering, etc. Please consult the [R-project homepage](#) for further information. **R-Forge** is the **new framework for R-project developers** based on [GFForge](#) offering easy access to the best in SVN, daily built and checked packages, mailing lists, bug tracking, message boards/forums, site hosting, permanent file archival, full backups, and total web-based administration.

### The new platform for the R Community

In order to get the most out of R-Forge, you'll need to [register as a site user](#). This will allow you to participate fully in all we have to offer. You may of course browse the site without registering, but will not have access to participate fully.

### Getting Your Package on R-Forge

[Register as a site user](#), then [Login](#) and finally, [Register Your Project](#). For details see the [documentation](#).

If you experience any problems or need help you can submit a [Support Request](#) to the [R-Forge Team](#) or write an email to [R-Forge@R-Project.org](mailto:R-Forge@R-Project.org).

Thanks... and enjoy the site.

### Latest News

#### depmixS4 version 0.2.0 now on CRAN

*Ingmar Visser - 2008-07-30 19:23 - depmixS4 - hidden Markov model classes*  
Most important changes include addition of more GLM distributions, addition of mix class to fit mixtures and latent class models, ie without dependencies, addition of a simulate function to generate new data from (fitted) models; finally, the help files have undergone major restructuring (as well as the internal structure of R files).  
(0 Comment) [\[Read More/Comment\]](#)

#### Compatibility

*Jose Claudio Faria - 2008-07-24 17:41 - TinnR*  
The TinnR package it is necessary to run TinnR version >= 2.0.0.1.

The version 2.0.0.1 pre-release was released today (jul/22/2008) to three users to tests.  
(0 Comment) [\[Read More/Comment\]](#)

#### eRm 0.10.1 on CRAN

*Patrick Mair - 2008-07-24 17:40 - eRm - extended Rasch modeling*  
The new eRm version provides two new functions:  
- `stepwisefit()` for automated stepwise term elimination according to different criteria  
- `goffRT()` contains various goodness-of-fit indices and model tests from logistic regression  
(0 Comment) [\[Read More/Comment\]](#)

#### sensitivity 1.4.0 on CRAN

*Gilles Pujol - 2008-07-21 12:14 - sensitivity*  
The release 1.4.0 of the package sensitivity is now on CRAN. Below are the details of changes compared to the last version.

### R-Forge Statistics

Hosted Projects: **195**  
Registered Users: **500**

### Most Active This Week

(100.0%) [mi-missing data imputation](#)  
(97.4%) [Rmetrics - Computational Finance](#)  
(94.7%) [ouch](#)  
(92.1%) [distr - S4 classes for distributions](#)  
(89.5%) [Roxysgen](#)  
(86.8%) [RobASt - Robust Asymptotic Statistics](#)  
(84.2%) [phylobase](#)  
(81.6%) [Matrix - Methods and Classes](#)  
(78.9%) [arm](#)  
(76.3%) [pomp](#)  
(73.7%) [R/GNU Linear Programming Kit Interface](#)  
(71.1%) [picante](#)  
(68.4%) [sts - extensible time series](#)  
(65.8%) [R GUI Generator \(RGG\)](#)  
(63.2%) [Classification and Regression Training](#)  
(60.5%) [PsychoR](#)  
(57.9%) [Integer and Nonlinear Optimization in R](#)  
(55.3%) [simacol - Ecological Models with R](#)  
(52.6%) [Data Analysis Using Regression](#)  
(50.0%) [TreEvo](#)

[\[ More \]](#)

### Recently Registered Projects

(08/01) [Phylogenetic Analysis Helper](#)  
(07/31) [BiplotGUI: Interactive Biplots in R](#)  
(07/30) [R4X](#)  
(07/30) [operators](#)  
(07/29) [Rumen digestion model](#)  
(07/29) [A R/C++ Template Library](#)  
(07/28) [SeqnR: Biological Sequences in R](#)  
(07/24) [Data Analysis Using Regression](#)  
(07/23) [TreEvo](#)  
(07/21) [TinnR](#)

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, ...



Provides the infrastructure for representing, manipulating and analyzing transaction data and patterns (frequent itemsets and association rules).

### Download

- Development Status: 5 - Production/Stable
- Intended Audience: Developers, End Users/Desktop
- License: GNU General Public License (GPL)
- Natural Language: English
- Operating System: MacOS, Windows, Linux
- Programming Language: C/C++, R
- Topic: Machine Learning

Registered: 2007-05-21 15:55

Activity Percentile: 0%

View project activity [statistics](#).

View list of [RSS feeds](#) available for this project

To install packages directly within R type `install.packages("packagename", repos="http://R-Forge.R-project.org")`

### Developer Info

Project Admins:  
 Michael Hahsler  
 Developers:  
 Christian Buchta  
 Kurt Hornik

[\[View Members\]](#)

[\[Request to join\]](#)

### Public Areas



[Project Home Page](#)



[Tracker](#)

- [Bugs](#) (0 open /3 total)

Bug Tracking System

- [Feature Requests](#) (0 open /0 total)

Feature Request Tracking System



[Public Forums](#) (4 messages in 2 forums)



[SCM Repository](#) (SVN: 170 updates, 11 adds)

### Latest News

#### arules supports now PMML

Michael Hahsler - 2008-05-02 10:52

(0 Comment) [\[Read More/Comment\]](#)

#### arules moved to R-Forge

Michael Hahsler - 2007-06-06 12:03

(0 Comment) [\[Read More/Comment\]](#)

[\[News archive\]](#)

[\[Submit News\]](#)

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



R Development Page | [Admin](#)

## 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.

### fArma - Rmetrics - ARMA Time Series Modelling

Environment for teaching "Financial Engineering and Computational Finance"

**Download:** [Package source \(.tar.gz\)](#) | [Windows binary \(.zip\)](#) | [MacOS X binary \(.tgz\)](#)

Linux x86_32	Linux x86_64	Windows x86_32	MacOS X universal
N/A	R build	R build	R build
R check	R check	R check	N/A

[Submit this package to CRAN]

To install the package directly within R type: `install.packages("fArma", repos="http://R-Forge.R-project.org")`

Version: 270.73 | Last change: 2008-07-23 17:14:26+02 | Rev.: 3407

### fAsianOptions - Rmetrics - EBM and Asian Option Valuation

Environment for teaching "Financial Engineering and Computational Finance"

**Download:** [Package source \(.tar.gz\)](#) | [Windows binary \(.zip\)](#) | [MacOS X binary \(.tgz\)](#)

Linux x86_32	Linux x86_64	Windows x86_32	MacOS X universal
N/A	R build	R build	R build
R check	R check	R check	N/A

[Submit this package to CRAN]

To install the package directly within R type: `install.packages("fAsianOptions", repos="http://R-Forge.R-project.org")`

Version: 270.73 | Last change: 2008-07-23 17:14:26+02 | Rev.: 3407

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





## 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 R CMD check, ...)
- ▶ Altering changelog messages in SVN repository
- ▶ Project renaming

## Open bugs:

- ▶ 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](mailto:R-Forge@R-project.org)  
Support tracker:  
[https://R-Forge.R-project.org/tracker/?group\\_id=34](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  
email: [Stefan.Theussl@wu-wien.ac.at](mailto:Stefan.Theussl@wu-wien.ac.at)  
URL: <http://statmath.wu-wien.ac.at/~theussl>

