

R: Software Development Life Cycle
A Description of R's Development, Testing, Release and
Maintenance Processes

April 4, 2014

The R Foundation for Statistical Computing
c/o Institute for Statistics and Mathematics
Wirtschaftsuniversität Wien
Welthandelsplatz 1
1020 Vienna, Austria
Tel: (+43 1) 31336 4754
Fax: (+43 1) 31336 904754
Email: R-foundation@R-project.org

Contents

1	The Scope of this Document	3
2	The R Foundation For Statistical Computing	5
3	What is R?	6
4	Software Development Life Cycle (SDLC)	8
4.1	Operational Overview	8
4.2	Source Code Management	8
4.3	Testing and Validation	9
4.4	Release Cycles	10
4.5	Availability of Current and Historical Archive Versions	11
4.6	Maintenance, Support and Retirement	11
4.7	Qualified Personnel	12
4.8	Physical and Logical Security	13
4.9	Disaster Recovery	13
5	Bibliography	14

Acknowledgements

The contributions of Marc Schwartz, Frank Harrell, Jr., Anthony Rossini and Ian Francis, who wrote and updated the initial versions of this document, are gratefully acknowledged.

1 The Scope of this Document

It is important to clarify that this document is SOLELY applicable to R software that is part of the official R distribution, as formally released by the R Foundation. This software is commonly referred to as “Base R plus Recommended Packages” and is released in both source code and binary executable forms under the [Free Software Foundation’s GNU Public License](#) (hereafter referred to as the GPL).

As of this writing, “Base R” includes the following packages:

- base
- compiler
- datasets
- graphics
- grDevices
- grid
- methods
- parallel
- splines
- stats
- stats4
- tcltk
- tools
- utils

and the “Recommended Packages” includes the following packages:

- boot
- class
- cluster
- codetools
- foreign
- KernSmooth
- lattice
- MASS
- Matrix
- mgcv

- nlme
- nnet
- rpart
- spatial
- survival

This document is NOT in any fashion, applicable to other R-related software and add-on packages made available via other parties, such as users or even members of the [R Development Core Team](#), who may, from time to time, make their software available via the [Comprehensive R Archive Network \(CRAN\)](#) or other software distribution repositories and vehicles.

It is important to note that there is a significant obligation on the part of the end-user's organization to define, create, implement and enforce R installation, validation and utilization related Standard Operating Procedures (SOPs) within the end-user's environment. These SOPs should define appropriate and reasonable quality control processes to manage end-user related risk within the applicable operating framework. The details and content of any such SOPs are beyond the scope of this document.

This document is not intended to be prescriptive, does not render a legal opinion and does not confer or impart any binding or other legal obligation. It should be utilized by the reader and his or her organization as one component in the process of making informed decisions as to how best to meet relevant obligations within their own professional working environment.

The R Foundation for Statistical Computing makes no warranties, expressed or implied, in this document.

The R Foundation For Statistical Computing

2 The R Foundation For Statistical Computing

The R Foundation is a not-for-profit organization working in the public interest. It was founded by the members of the R Development Core Team in order to:

- Provide support for the R project and other innovations in statistical computing. We believe that R has become a mature and valuable tool and we would like to ensure its continued development and the development of future innovations in software for statistical and computational research.
- Provide a reference point for individuals, institutions or commercial enterprises that want to support or interact with the R development community.
- Hold and administer the copyright of R software and documentation.

R is an official part of the [Free Software Foundation's](#) GNU project, and the R Foundation has similar goals to other open source software foundations, such as the [Apache Foundation](#) and the [GNOME Foundation](#).

Among the goals of the R Foundation are the support of continued development of R, the exploration of new methodology, teaching and training for statistical computing and the organization of meetings and conferences with a statistical computing orientation.

The R Foundation is seated in Vienna, Austria and currently hosted by the Vienna University of Technology. It is a registered association under Austrian law and active worldwide. The R Foundation can be contacted at:

The R Foundation for Statistical Computing
c/o Institute for Statistics and Mathematics
Wirtschaftsuniversität Wien
Welthandelsplatz 1
1020 Vienna, Austria
Tel: (+43 1) 31336 4754
Fax: (+43 1) 31336 904754
Email: R-foundation@R-project.org

The R Foundation Statutes are available from the Foundation's web site:

<http://www.r-project.org/foundation/>

3 What is R?

Introduction to R

R is a language and environment for statistical computing and graphics. It is a GNU project and is similar to the S language and environment that was developed at Bell Laboratories (formerly AT&T, now Lucent Technologies) by John Chambers and his colleagues². R can be considered as a distinct implementation of S, developed separately from the original implementation at Bell Laboratories. Although there are some important differences between these two implementations, much code written for S runs unaltered under R.

R provides a wide variety of statistical (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering, etc.) and graphical techniques, and is readily extensible. The S language is often the vehicle of choice for research in statistical methodology and R provides an open source route to participation in that activity.

One of R's strengths is the ease with which well designed publication-quality plots can be produced, including mathematical symbols and formulae where needed. Great care has been taken over the defaults for the minor design choices in graphics, but the user retains full control.

R is available as Free Software under the terms of the Free Software Foundation's GNU General Public License in source code form. It compiles and runs on a wide variety of UNIX platforms and similar systems (including FreeBSD and Linux), Windows and MacOS.

The R environment

R is an integrated suite of software facilities for data manipulation, calculation and graphical display. It includes:

- an effective data handling and storage facility,
- a suite of operators for calculations on arrays, in particular matrices,
- a large, coherent, integrated collection of intermediate tools for data analysis,
- graphical facilities for data analysis and display either on-screen or on hardcopy, and
- a well developed, simple and effective programming language that includes conditionals, loops, user-defined recursive functions and input and output facilities.

The term “environment” is intended to characterize R as a fully planned and coherent system, rather than an incremental accretion of very specific and inflexible tools, as is frequently the case with other data analysis software.

R, like S, is designed around a true programming language, and it allows users to add additional functionality by defining new functions. Much of the system is itself written in the R dialect of the S language, which makes it easy for users to follow the algorithmic choices made. For computationally intensive tasks, C, C++ and Fortran code can be linked and called at run time. Advanced users can write C code to manipulate R objects directly.

Many users think of R as a statistics system. We prefer to think of it of an environment within which statistical techniques are implemented. R can be extended (easily) via packages. There are a number of

² See references [Becker et al.(1988)Becker, Chambers, and Wilks], [Chambers and Hastie(1992)] and [Chambers(1998)]

packages listed previously supplied with the R distribution and many more, covering a very wide range of modern statistics, are available through the CRAN family of Internet sites.

R has its own L^AT_EX-like documentation format, which is used to supply comprehensive documentation, both on-line in a number of formats and in hardcopy.

In addition, as R is open source, the availability of R's source code provides for superior and thorough documentation of R's functionality and designed behavior and is open to inspection by all users.

4 Software Development Life Cycle (SDLC)

4.1 Operational Overview

The development, release and maintenance of R is, broadly, a collaborative process involving the [R Development Core Team](#) (hereafter referred to as R Core). Members of R Core represent multiple statistical disciplines and are based at academic, not-for-profit and industry-affiliated institutions on multiple continents.

Most communications amongst the members of R Core take place electronically via e-mail and similar means. A non-public e-mail list ([r-core](#)) provides a common forum for discussions amongst the members of R Core. An archive of the list is available to facilitate R Core in documenting and reviewing these discussions, as they pertain to development decisions and related issues.

R Core does meet, collectively and/or in smaller groups, with a level of frequency dictated by multiple factors, including taking advantage of regularly scheduled conferences where members of R Core may already be in attendance. Such conferences include those that are specific to statistical computing and R itself (<http://www.r-project.org/conferences.html>). These routine communications and meetings ensure that the collaborative efforts are appropriately coordinated and prioritized as ongoing development takes place.

Reasonable software development and testing methodologies are employed by R Core in order to maximize the accuracy, reliability and consistency of R's performance. While some aspects of R's development are handled collaboratively, others are handled by members of the team with specific interests and expertise in focused areas.

Importantly, as R is released under the terms of the GPL, all of the source code underlying R, whether it be in R, C or FORTRAN, is available for peer review by all members of the R user community. Thus, all of the functionality embodied within R is subject to continuous critique and improvement relative to its accuracy, reliability and consistency.

The size of the R user community (difficult to define precisely, because there are no sales transactions, but conservatively estimated as being in the tens of thousands, with some independent estimates in the hundreds of thousands), provides for extensive review of source code and testing in “real world” settings outside the confines of the formalized testing performed by R Core. This is a key distinction, related to product quality, between R and similar software that is only available to end users in a binary, executable format. In conjunction with detailed documentation and references provided to end users, the size of the R user community, all having full access to the source code, enables a superior ability to anticipate and verify R's performance and the results produced by R.

Additional documentation regarding the activities of R Core as they pertain to development, goals and related activities, including coding guidelines, are available for review:

- [R Developer Page](http://developer.r-project.org/) (<http://developer.r-project.org/>)
- [R Internals – A Guide to the Internal Structures of R and Coding Standards for the R Core Team](http://cran.r-project.org/doc/manuals/R-ints.html) (<http://cran.r-project.org/doc/manuals/R-ints.html>)

4.2 Source Code Management

All of R's source code is managed in a source code version control repository based on [Subversion](#). The [R Subversion Repository](#) is access controlled, such that only members of R Core have write access to the

source code tree. Various security, access control and archival procedures are in place to provide reasonable protection and to maintain the integrity of the hosting server and the source code management system.

Separate source code branches for version control are maintained by R Core. The current Release Branch and the ongoing Development Version are kept in separate branches to facilitate non-conflicting source code management. The Release Branch is designed for bug fixes and allows only minor feature enhancements. Major features are introduced in the Development Version, from which a new Release Branch is made prior to the next x.y.0 release.

Daily logs of code changes are maintained within the Subversion repository and reflect all aspects of code changes made by R Core. These logs are available for public review as http://developer.r-project.org/R_svnlog_YYYY, where 'YYYY' is a placeholder for a four-digit year specification (e.g. 2011).

In addition, a "NEWS" file is actively maintained by R Core to make it easier for users to track changes made to past, present and future versions of R. The current version of this file is available for public viewing at <http://stat.ethz.ch/R-manual/R-devel/doc/html/NEWS.html>. This file is also included in all source code and binary executable versions of R to enable end users to review and gain insight into the ongoing changes to R via the `news()` function.

The typical format of the NEWS file contains detailed, version-specific information on:

- Significant User-Visible Changes
- New Features
- Graphics Devices
- Installation
- Package Installation
- Utilities
- Deprecated and Defunct
- C-Level Facilities
- Bug Fixes

The entire list (and any additions) may or may not be present for each R version as appropriate.

Further, older versions of the NEWS file are available as <https://svn.r-project.org/R/trunk/ONEWS> and <https://svn.r-project.org/R/trunk/OONEWS>. These files enable R users to gain insight into the full history of R's ongoing development, back to version 0.50, which was released in 1997.

4.3 Testing and Validation

Within the R Core development related documents, as identified in the aforementioned references (see Section 4.1), guidelines are provided relative to modifications to source code, regression tests, validation tests and similar issues. These guidelines are in place to maximize code quality and to facilitate ongoing code validation during development and during the "run-up" to each version release.

A set of validation tests are maintained and upgraded by R Core to enable the testing of source code against known data and known results. Any errors noted during this testing are resolved prior to release.

The tests are located in the “tests” sub-directory of the extracted source code tarball. A README file is also available in that directory to describe the procedures to run the tests and various options related to selecting all tests or only a subset of the tests to run. The source code and expected results for these tests are available for review and use in other applications as may be appropriate.

These tests are also available to end users and/or system administrators and can be run as part of their installation process to provide further documentation and objective evidence as to the accuracy, reliability and consistency of their installation of R.

As with any statistical software, the user should take care to consider the appropriateness of any R software, and the statistical methods implemented in the software, to the intended application. The potential exists in any statistical software for the lack of consistency and reliability in results due to the inappropriate application of statistical methodologies. Reasonable judgment in this regard should be rendered by users with appropriate expertise.

The entire R source code tree is available to end users (either via the Subversion repository or via source code archive files, known as “tarballs”, that are automatically created with daily updates). Additional testing is solicited from the user community during so-called “Alpha”, “Beta” and “Release Candidate” testing cycles. Progressively stronger restrictions are imposed on modifications to the source code during the testing cycles to minimize the risk of unexpected side effects. This provides further opportunities to identify and resolve issues that may have been missed during the development process, such as “boundary” issues that may represent unusual or atypical circumstances, including unique operating system and/or hardware configurations.

Feedback from the community is facilitated by the use of the [r-devel e-mail list](https://stat.ethz.ch/mailman/listinfo/r-devel)³ and via the [R Bug Tracking System](http://bugs.r-project.org/)⁴. This open and public process enables a wider array of code testing and further increases the likelihood of resolving issues prior to the release of a stable version of R.

4.4 Release Cycles

Once the in-development version of R has been approved for release by R Core’s designated Release Manager, a public announcement is made via the R e-mail lists to the user community.

Source code archive files (“tarballs”) are made available via the CRAN mirror infrastructure.

Pre-built executable binary install files follow and are made available for common operating system and CPU architectures. These include Linux, Windows and MacOS platforms.

R’s major release cycles are generally predictable. Prior to R version 3.0.0, x.y.0 releases occurred on or about April 1 and October 1 of each calendar year. Effective with R version 3.0.0, x.y.0 releases will occur on or about April 1 of each calendar year.

x.y.z versions, so called patch releases are made available when required in order to fix issues discovered in the current release. Generally one or two of these releases have been made before the semi-annual release of the next x.y.0 version and this number may increase with the introduction of the annual release cycle.

Additional instructions regarding the utilization of R source code, installation requirements, compilation and platform and operating system related issues are extensively documented in the R Installation and Administration Manual, which is available with source code and binary executables and online at <http://cran.r-project.org/manuals.html>.

³ <https://stat.ethz.ch/mailman/listinfo/r-devel>

⁴ <http://bugs.r-project.org/>

4.5 Availability of Current and Historical Archive Versions

Current and historical versions of R are available in source code tarballs from the [main CRAN server](http://cran.r-project.org/src/base/) (<http://cran.r-project.org/src/base/>) and its [worldwide mirrors](http://cran.r-project.org/mirrors.html) (<http://cran.r-project.org/mirrors.html>).

Pre-built executable binary install files for current versions are made available for common operating system and CPU architectures. These include Linux, Windows and MacOS platforms and are available from the [main CRAN server binary tree](http://cran.r-project.org/bin/) (<http://cran.r-project.org/bin/>) and from its worldwide mirrors as referenced above.

4.6 Maintenance, Support and Retirement

Each Released Version of R is actively supported by R Core with respect to bug reporting, fixes and patches. Patched versions are made available, generally as source code only, to end users to facilitate their installation of these. Binary executable installation files for the patched Release Versions are made available at the discretion of the individual maintainers of the platform specific versions.

Source code tarballs of the daily incremental patched versions of each current R release are made available to the community via an [FTP server](ftp://ftp.stat.math.ethz.ch/Software/R/) (<ftp://ftp.stat.math.ethz.ch/Software/R/>) for download to enable R users to update their systems between formal releases as their local needs may dictate.

In addition, users with Subversion clients can download the latest copy of the source code tree at any time, via a direct connection to the Subversion server.

As each version of R is released, there are a variety of support resources that are made available to the community of end users.

Extensive documentation is provided by R Core and is available both within the source code and binary executable versions of R as well as online in HTML and PDF formats at <http://cran.r-project.org/manuals.html>.

Function-specific help is also available within R including, where appropriate, extensive references to algorithms and methods to facilitate the user's comprehension of R's functionality and expected behavior.

R FAQs (Frequently Asked Questions) are also available to facilitate answers to commonly asked end user questions. These are available at:

- [The Main R FAQ](http://cran.r-project.org/doc/FAQ/R-FAQ.html) (<http://cran.r-project.org/doc/FAQ/R-FAQ.html>)
- [R FAQ for Windows](http://cran.r-project.org/bin/windows/base/rw-FAQ.html) (<http://cran.r-project.org/bin/windows/base/rw-FAQ.html>)
- [R FAQ for MacOS](http://cran.r-project.org/bin/macosx/RMacOSX-FAQ.html) (<http://cran.r-project.org/bin/macosx/RMacOSX-FAQ.html>)

R's Bug Reporting system, available online at <http://bugs.r-project.org/>, facilitates end user reporting of bugs identified during the course of use. In addition, an internal R function, `bug.report()`, is available to enable end users to generate and send bug reports directly from an interactive R session.

An extensive set of public e-mail lists exist. These are the primary vehicle for interactive support and communications between R Core and the user community. There are two primary lists, called `r-devel` and `r-help`.

The former list is principally for issues surrounding R's development and lower level coding issues that are more technical.

The latter list, which is the primary end user support forum, is an active discussion on various R coding or usage issues and related concerns.

Additional e-mail lists focus on specific special interest areas that range from database interfaces to robust statistics and financial modeling.

More information on these e-mail lists is available at <http://www.r-project.org/mail.html>.

Extensive search facilities, accessible at <http://www.r-project.org/search.html>, are also available to search the list archives, enabling users to perform keyword-based searches of prior discussions and the online documentation. An internal R function, `RSiteSearch()`, is also available to facilitate such searches during an interactive R session.

The R Journal (ISSN 2073-4859), formerly R News, a peer-reviewed newsletter is available electronically as a periodical from <http://journal.r-project.org/>. The R Journal provides general information on R and R Core and user contributed articles in specific domains of interest.

A large set of published books, several by members of R Core, are available to support the use of R, both generally and within subject-matter-specific domains. A periodically updated but partial list of these books is available at <http://www.r-project.org/doc/bib/R-books.html>.

The x.y.0 releases are maintained via a series of x.y.z patch releases. At a new x.y.0 version of R, the prior version is retired from formal support. R Core's efforts are then focused on the new Release (and the ongoing Development) version. No further development, bug fixes or patches are made available for the retired versions. Thus there is always only one current version of R. However, the SVN repository will allow older release branches to be reopened, should the need arise.

4.7 Qualified Personnel

As noted in Section 4.1, members of R Core represent multiple statistical disciplines and are based at academic, not-for-profit and industry-affiliated institutions on multiple continents.

All members of R Core hold Ph.D. and/or Master's degrees (all but one have Ph.D.s) from accredited academic institutions and have published extensively in peer reviewed journals. Several have written books on statistical computing technologies and applications. The members of R Core constitute a widely recognized, international team of experts on statistical computing and software development.

Institutions at which the members of R Core currently hold or have previously held appointments include:

- University of Wisconsin – Madison
- Bell Laboratories
- University of Copenhagen
- Fred Hutchinson Cancer Research Center
- Wirtschaftsuniversität Wien
- Università degli Studi di Milano
- University of Auckland
- Ludwig-Maximilians-Universität – München

- Technische Universität Dortmund
- University of Washington
- Eidgenössische Technische Hochschule Zürich
- University of Western Ontario
- Centre International de Recherche sur le Cancer, Lyon
- Oxford University
- Indian Statistical Institute, Delhi Centre
- University of California – Davis
- University of Iowa
- AT&T Research Labs

4.8 Physical and Logical Security

The R Foundation maintains its key servers within the brick and mortar infrastructure of university-supported computing facilities. In accordance with defined security policies, only personnel with authorized access may enter.

User names and passwords are required by all R Core members to gain access to computing systems for R Foundation-related activities. User accounts are limited in access based upon standard security policies and functional requirements.

Network access is controlled via the use of typical hardware and software controls, including the use of firewalls, security policies and related mechanisms.

4.9 Disaster Recovery

As a result of having R Foundation servers within the confines of university-hosted computing facilities, disaster recovery plans for R Foundation computing systems are in sync with those of the host facilities.

In addition, the worldwide network of CRAN mirrors provides for an alternative means of accessing key components of R, should primary servers be temporarily unavailable.

5 Bibliography

References

- [Becker et al.(1988)Becker, Chambers, and Wilks] R. A. Becker, J. M. Chambers, and A. R. Wilks. *The New S Language*. Chapman & Hall, London, 1988. ISBN 0-534-09193-8.
- [Chambers(1998)] J. M. Chambers. *Programming with Data*. Springer, New York, 1998. URL <http://cm.bell-labs.com/cm/ms/departments/sia/Sbook/>. ISBN 0-387-98503-4.
- [Chambers and Hastie(1992)] J. M. Chambers and T. J. Hastie. *Statistical Models in S*. Chapman & Hall, London, 1992. ISBN 0-534-16764-0.