

## **RStudio: Regulatory Compliance and Validation Issues**

### **A Guidance Document for the Use of RStudio Integrated Development Environment (IDE) Commercial Products in Regulated Clinical Trial Environments**

June 19, 2014

RStudio, Inc.

250 Northern Ave.

Boston, MA USA 02210

Tel: (+1) 844-448-1212

Email: [info@rstudio.com](mailto:info@rstudio.com)

# Regulatory Compliance and Validation Issues

## *A Guidance Document for the Use of the RStudio Interactive Development Environment Commercial Products in Regulated Clinical Trial Environments*

### **Purpose and Introduction**

The purpose of this document is to demonstrate that the RStudio Interactive Development Environment (IDE), when used in a qualified fashion, can support the appropriate regulatory requirements for validated systems, thus ensuring that resulting electronic records are “trustworthy, reliable and generally equivalent to paper records.”

**This document applies to RStudio IDE desktop and server products released in binary executable forms under an RStudio commercial license.**

This document is NOT in any fashion, applicable to other R-related software and add-on packages. 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). 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.

**RStudio, Inc. makes no warranties, expressed or implied, in this document.**

### **Validation of Systems for 21 CFR Part 11 Compliance**

Validation is defined by the FDA<sup>1</sup> as: “Establishing documented evidence which provides a high degree of assurance that a specific process will consistently produce a product meeting its predetermined specifications and quality attributes.”<sup>2</sup>

It is crucial to note that many validation requirements, as described in the following pages, may be met by the operational characteristics of software systems (i.e. operating systems and database applications) and other technologies or processes outside of the RStudio IDE itself, where RStudio IDE will be used as a component in an overall data management, analysis and presentation process.

---

<sup>1</sup> General Principles of Software Validation; Final Guidance for Industry and FDA Staff

<sup>2</sup> Glossary Of Computerized System and Software Development Terminology

# Software Development Life Cycle (SDLC)

## Operational Overview

The development, release and maintenance of RStudio IDE commercial products is a collaborative process.

Most communications among RStudio development team members take place electronically via e-mail and similar means. A non-public e-mail list provides a common forum for discussions along with video conferencing, instant messaging and group chat tools such as Hipchat.

RStudio development team members meet, collectively and/or in smaller groups, with a level of frequency dictated by multiple factors, including regularly scheduled company meetings. 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 in order to maximize the accuracy, reliability and consistency of RStudio IDE's performance. While some aspects of development are handled collaboratively, others are handled by members of the team with specific interests and expertise in focused areas.

Importantly, as RStudio IDE commercial products incorporate RStudio IDE open source components developed by the same team and released under the terms of the Affero General Public License (AGPL), much of the source code underlying RStudio IDE products is available for peer review by all members of the R user community. Thus, much of the functionality embodied within RStudio IDE commercial products is subject to continuous critique and improvement relative to its accuracy, reliability and consistency.

Additional documentation regarding RStudio IDE development activities as they pertain to development, goals and related activities, including coding guidelines, are available for review:

<https://github.com/rstudio/rstudio>

<https://github.com/rstudio/rstudio/wiki/RStudio-Development>

<https://github.com/rstudio/rstudio/wiki/Installing-RStudio-Dependencies>

## Source Code Management

All of RStudio IDE's commercial source code is managed in [git](#), a source code revision control software. The RStudio IDE Repository is access controlled, such that only members of RStudio development team have write access to the source code tree.

Separate source code branches for version control are maintained by RStudio developers. 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 release.

Daily logs of code changes are maintained within the repository and reflect all aspects of code changes.

In addition the development team tracks the release history here <https://support.rstudio.com/hc/en-us/articles/200716783-RStudio-Release-History> and the associated release notes here [http://www.rstudio.com/ide/docs/release\\_notes\\_v0.98](http://www.rstudio.com/ide/docs/release_notes_v0.98) .

Links to release notes for previous versions of the software are maintained at the end of the current release note document.

The typical format of the release notes includes:

### Feature Area

- New additions

### RStudio Server

- New additions

### Misc.

- New additions

### Bug fixes

- Bug descriptions

## Testing and Validation

Within the RStudio development team, 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 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 `src/gwt/test` directory of the source code control system.

In general, regressions not caught in formal validation tests are almost all caught within a few days in community testing.

Core software code included in daily builds of RStudio’s open source IDE is downloaded on a daily basis and tested by thousands of users. Additional testing of RStudio commercial IDE products is solicited from the user community during so-called “Alpha” and “Beta” preview releases, several months ahead of general availability of a major release.

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.

Feedback from the community is facilitated by the use of [support.rstudio.com](https://support.rstudio.com) where users report issues and seek support. This process enables a wider array of code testing and further increases the likelihood of resolving issues prior to the release of a stable version.

## Release Cycles

Once the in-development version of RStudio IDE commercial products have been approved for release by a designated Release Manager, a public announcement is made via the [RStudio blog](#).

Source code archive files for the open source components are made available at <https://github.com/rstudio/rstudio>

Pre-built executable binary install files for RStudio IDE commercial products follow on RStudio.com and are made available for common operating system and CPU architectures. These can include Linux, Windows and MacOS platforms.

Patch releases are made available when required in order to fix issues discovered in the current release.

Additional instructions regarding the use of RStudio IDE commercial products, installation requirements and platform and operating system related issues are extensively documented in the RStudio IDE Server Pro Administration Guide, which is available online here

<https://support.rstudio.com/hc/en-us/categories/200035113-Documentation> and here <https://s3.amazonaws.com/rstudio-server/rstudio-server-pro-0.98.501-admin-guide.pdf>.

## **Availability of Current and Historical Archive Versions**

Source code for every open source IDE version we've ever released is available via (note the version number in the URL):

<https://github.com/rstudio/rstudio/tree/v0.98.501>

Binaries for every version we've ever released are available via (note the version number in the URL). For example, the desktop IDE version is available here::

<https://s3.amazonaws.com/rstudio-dailybuilds/RStudio-0.98.501.exe>

The examples above assume v0.98.501 -- you can substitute any other version to get the source or binary.

## **Maintenance, Support and Retirement**

Each Released Version of an RStudio IDE commercial product is actively supported by RStudio, Inc. with respect to bug reporting, fixes and patches. Binary executable installation files for patched Release Versions are made available at the discretion of RStudio.

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

Extensive documentation is provided by RStudio in HTML and PDF formats at

<http://www.rstudio.com/ide/docs/> and <https://s3.amazonaws.com/rstudio-server/rstudio-server-pro-0.98.501-admin-guide.pdf>

A set of published books by members of the RStudio team are available to support the use of R and RStudio products. A periodically updated but partial list of these books is available at <http://www.rstudio.com/training/books.html>

## **Qualified Personnel**

All members of RStudio's development team hold qualifying degrees and prior development experience, many with Ph.D. and/or Master's degrees from accredited academic institutions. Many have published in peer reviewed journals. Several have written books on statistical computing technologies and applications. The members of RStudio's development team constitute a widely recognized, international team of experts on statistical computing and software development.

Institutions at which the members of RStudio development team members currently hold appointments or have previously been affiliated include:

University of California - Davis

Harvard University

Iowa State University

Macalester College

Massachusetts Institute of Technology

University of Massachusetts (Amherst)

Northeastern University

Northwestern University

Rice University

## **Physical and Logical Security**

RStudio, Inc. maintains its key servers with Amazon. Secure Shell (SSH) private keys protect access in accordance with Amazon's defined security policies.

Amazon requires user names and passwords for all RStudio development team members to gain access to computing systems for RStudio-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.

## **Disaster Recovery**

RStudio commercial IDEs are installed on customer systems and therefore, subject to customer disaster recovery practices. RStudio does not operate IDE servers that would require continuous uptime.

For delivery of RStudio commercial IDE product purchases and upgrades, we rely on Amazon for availability of our binaries and disaster recovery practices. For development of RStudio commercial IDE products the RStudio development team relies on Github for the availability of source code and disaster recovery practices. However, because of the distributed nature of our development process and infrastructure, even in the event of an unrecoverable disaster impacting either service, RStudio could continue to make binary products available to customers and source products available to our development team.

## **21 CFR Part 11 Compliance Functionality**

### **Overview**

Within the regulated domain, RStudio IDE commercial products are intended to be utilized as a component within a larger data management framework, with respect to data acquisition, validation and related source electronic *records* tasks. RStudio IDE commercial products' design and development are focused on statistical application construction, predominantly using the R programming language, rather than on data management tasks such as transaction/data processing and related functionality.

To that end, the following sections discuss important components of the 21 CFR Part 11 Regulation, provides RStudio's interpretation of each, and discuss how RStudio IDE commercial products and other enabling technologies, within an overall data management framework, can meet the guidance interpretations.

Note that sections 11.10(a) and (i), pertaining to system validation and qualified personnel, respectively, have already been covered previously.

In the following sections, the term *record* means an electronic record that is interpreted to fall within the remit of Part 11 as defined in FDA Guidance for Industry Part 11, Electronic Records; Electronic Signatures – Scope and Application (2003).

**RStudio IDE commercial products are not intended to create, maintain, modify or delete Part 11 relevant *records* but to perform calculations and draw graphics.**

**11.10(b) The ability to generate accurate and complete copies of records in both human readable and electronic form suitable for inspection, review, and copying**

RStudio understands this item to mean that any *records* created or maintained in the system must be accurate and complete. These *records* must be available in both human readable and electronic form.

RStudio IDE commercial products are not intended to create, maintain, modify or delete Part 11 relevant *records* but to perform calculations and draw graphics using the R statistical programming language. Where RStudio IDE commercial products' use may be interpreted as creating *records*, however any such *records* (for example data objects such as vectors, matrices, lists and data frames, and graphics, plots and images) are available to be output in various industry-standard formats. Because R provides for the routine generation of these outputs as standard features, the output is available in both machine- and human-readable formats.

Using these industry-standard formats, the output is available to be read by other products that also utilize these same industry standards and these *records* are therefore readable independent of the use of R and RStudio IDE commercial products.

In conjunction with local policies regarding *record* access control, retention and archival, RStudio IDE commercial products meet the FDA requirements for the inspection, review and copying of *records* as defined above.

**11.10(c) Protection of records to enable their accurate and ready retrieval throughout the records retention period**

RStudio, Inc. understands this item to mean that all *records* created or maintained in RStudio IDE commercial products must be stored in a manner that enables accurate and ready retrieval.

RStudio IDE commercial products are not intended to create, maintain, modify or delete Part 11 relevant *records* but to perform calculations and draw graphics.

Therefore, *Records* created by RStudio IDE commercial products will, therefore, reside within and be managed by a separate host system.

The host system is required to provide for compliance with this part using local policies regarding the retention and archival of such *records* and the mechanisms and access controls in place.

#### **11.10(d) Limiting system access to authorized individuals**

RStudio, Inc. understands this item to mean that access to the computer system that creates, maintains or modifies a *record* is limited to only authorized individuals.

RStudio IDE commercial products are applications that run within the hosting computer environment, which must provide user access controls at hardware and/or operating system levels.

The requirement for this section is typically met via system level functionality and is based on user roles, object level security and related security policies.

Approved users must be supplied unique user account identifiers and passwords, which are required to gain access to the hosting systems and thus to RStudio IDE commercial products. Upon connection to the hosting system, RStudio IDE Server Pro - but not RStudio IDE Commercial Desktop License - provides further access and functional restrictions to limit the activities in which the user may engage.

#### **11.10(e) Use of secure, computer-generated, time-stamped audit trails to independently record the data and time of operator entries and actions that create, modify, or delete electronic records. Record changes shall not obscure previously recorded information. Such audit trail documentation shall be retained for a period at least as long as that required for the subject electronic records and shall be available for agency review and copying**

RStudio, Inc. understands this item to mean that the creation, modification or deletion of *records* must have an associated audit trail describing who, when and why an action was performed. Additionally, any such audit trail will be also considered an electronic *record* within the scope of Part 11.

RStudio IDE commercial products are not intended to create, maintain, modify or delete Part 11 relevant *records* but to perform calculations and draw graphics.

Where RStudio IDE commercial products' use may be interpreted as creating *records*, however, its use of the language R can support audit trail creation within the *record*.

R includes `date()`, `Sys.time()`, `Sys.Date()` and `Sys.timezone()` functions which enable users to include date and time stamps on report, graphical and other output, thus enabling the use of this information in the tracking of user sessions.

*Records* created by RStudio IDE commercial products necessarily reside within and are managed by a separate host system. Therefore, after *record* creation, any subsequent changes to the *record* must have an audit history imposed by the host system. This may be implemented technically via system-level logging as a component of the hosting computer system.

For session-based logging focusing on data analysis, the organization using RStudio IDE commercial products would need to provide extensions using R or other tools to facilitate the generation of a session-based audit trail that meets the local implementation requirements of the organization's quality assurance group. The security and integrity of this log would be ensured through the use of the hosting system's user and object-based security models.

#### **11.10(f) Use of operational system checks to enforce permitted sequencing of steps and events, as appropriate**

RStudio, Inc. understands this item to mean that effective user technology, processes, and interfaces must be in place to reduce errors made by an operator to the extent that system errors can be minimized.

RStudio IDE commercial products were designed with an architecture, technology, process and interface that provide operator flexibility. Aside from enabling visual confirmation of code execution within the interactive development environment, RStudio makes no checks for software function or features.

These capabilities are similar to those of any statistical software IDE.

Appropriate coding techniques that implement good and defensive programming style are documented and described in many books, including *Software for Data Analysis* (Chambers)<sup>[2]</sup>.

**11.10(g) Use of authority checks to ensure that only authorized individuals can use the system, electronically sign a record, access the operation or computer system input or output device, alter a record, or perform the operation at hand**

RStudio, Inc. understands this item to mean that the system must provide for authority checks to allow users to perform system operations, such as applying electronic signatures, access to input and output devices, the ability to alter a *record* and perform functions.

Authority checks (such as user name/password controls) must be implemented within the host system, as described in section 11.10(d) and can be further enhanced by using features in RStudio IDE Server Pro. This provides for controlled access for authorized users to the host system and to the RStudio IDE Server Pro commercial application.

Within RStudio IDE Server Pro there are controls to enable/disable user access. Any additional restrictive functionality must be implemented within the host system according to relevant business processes and documented operating procedures as defined by the user's organization.

**11.10(h) Use of device (e.g., terminal) checks to determine, as appropriate, the validity of the source of data input or operational instruction**

RStudio, Inc. understands that these checks are warranted where only certain devices have been selected as legitimate sources of data input or commands. The device checks would be used to determine if the data or command source was authorized. If RStudio IDE commercial products are used as a primary-source data management and data entry system, such checks would need to be implemented by the developer of the code.

RStudio IDE Server Pro enhances the host environment capabilities as discussed previously, notably in sections 11.10(d) and 11.10(f).

RStudio IDE commercial products are not intended to create, maintain, modify or delete Part 11 relevant *records* but to perform calculations and draw graphics.

**11.10(j) The establishment of, and adherence to, written policies that hold individuals accountable and responsible for actions initiated under their electronic signatures, in order to deter record and signature falsification**

RStudio, Inc. understands that individuals must understand their responsibility and accountability when performing actions using their electronic signatures. This must be communicated with documented policies.

RStudio IDE commercial products are not intended to create *records* but to perform calculations and draw graphics. Following from this, they are not intended to allow for signature of *records*.

**11.10(k) Use of appropriate controls over systems documentation**

21 CFR Part 11.10(k) indicates that these controls must include:

Adequate controls over the distribution of, access to, and use of documentation for system operation and maintenance

Revision and change control procedures to maintain an audit trail that documents time-sequenced development and modification of systems documentation

RStudio, Inc. understands this item to mean that there must be control over who can access and change system documentation and also that there exists revision and change control in place for system documentation.

All releases of RStudio IDE commercial products include documentation covering installation, administration, programming and related user guides. Documentation is created once per Release Version; thus these documents are uniquely identifiable and associated with a specific release of the software.

This documentation is published and maintained by RStudio as part of the Software Development Life Cycle using the [Git](#) version-control system. This documentation is controlled in the same manner as RStudio IDE commercial products' source code.

This documentation is provided to RStudio IDE commercial products users in electronic formats.

The maintenance and distribution of this documentation at the RStudio user site is the sole responsibility of RStudio, Inc. and is handled in accordance with training and other standard operational procedures.

**Section 11.30 Controls for Open Systems - the system shall employ procedures and controls designed to ensure the authenticity, integrity and as appropriate the confidentiality of electronic records from the point of their creation to the point of their receipt. Additional measures such as document encryption and use of appropriate digital signature standards to ensure, as necessary under the circumstances record authenticity, integrity and confidentiality**

RStudio IDE commercial products support the host environment (see previous discussion, particularly section 11.10(d)) that provides these capabilities.

It is the sole responsibility of the RStudio IDE commercial products' user to ensure that the appropriate safeguards are implemented for a particular hosting system.

## 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.

---

[1]  
<https://stat.ethz.ch/mailman/listinfo/r-devel> <sup>6</sup>  
<http://bugs.r-project.org/>

[2] See reference [[Chambers\(2008\)](#)]