

RStudio: Regulatory Compliance and Validation Issues

A Guidance Document for the Use of RStudio Professional Products in Regulated Clinical Trial Environments

June 26, 2019

RStudio, Inc.
250 Northern Ave.
Boston, MA USA 02210

Tel: (+1) 844-448-1212
Email: info@rstudio.com

Contents

1	Purpose and Introduction	1
2	Validation of Systems for 21 CFR Part 11 Compliance	1
3	Software Development Life Cycle (SDLC)	2
3.1	Operational Overview	2
3.2	Source Code Management	2
3.3	Testing and Validation	2
3.4	Release Cycle	3
3.5	Availability of Current and Historical Archive Version	3
3.6	Maintenance, Support, and Retirement	3
3.7	Qualified Personnel	4
3.8	Physical and Logical Security	4
3.9	Disaster Recovery	5
4	21 CFR Part 11 Compliance Functionality	5
4.1	Overview	5
4.2	11.10(b) The ability to generate accurate and complete copies of <i>records</i> in both human readable and electronic form suitable for inspection, review, and copying	6
4.3	11.10(c) Protection of records to enable their accurate and ready retrieval throughout the records retention period	6
4.4	11.10(d) Limiting system access to authorized individuals	6
4.5	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	7
4.6	11.10(f) Use of operational system checks to enforce permitted sequencing of steps and events, as appropriate	7
4.7	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	8
4.8	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	8

4.9	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	8
4.10	11.10(k) Use of appropriate controls over systems documentation	9
4.11	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	9

5 Bibliography **10**

Contents

1 Purpose and Introduction

The purpose of this document is to demonstrate that RStudio Professional Products, 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 Professional Products released in binary executable forms under an RStudio commercial license. RStudio currently releases five such products: RStudio Server Pro, RStudio Desktop Pro, RStudio Connect, RStudio Package Manager, and Shiny Server Pro.

RStudio Server Pro and **RStudio Desktop Pro** are Integrated Development Environments (IDEs) for statistical programming, primarily in the R programming language.

RStudio Connect and **Shiny Server Pro** are server-based software to allow the sharing and publishing of data artifacts. They also provide tools for user access and management and helping users scale and manage their R processes.

RStudio Package Manager is server-based software to help IT departments manage add-on R packages. For the duration of this document, these products will collectively be referred to as RStudio Professional Products.

This document is NOT in any fashion, applicable to any other R-related software or 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.

2 Validation of Systems for 21 CFR Part 11 Compliance

Validation is defined by the FDA¹ 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.”²

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 RStudio Professional Products, where RStudio Professional Products will be used as components in an overall data management, analysis, and presentation process.

¹General Principles of Software Validation; Final Guidance for Industry and FDA Staff

²Glossary Of Computerized System and Software Development Terminology

3 Software Development Life Cycle (SDLC)

3.1 Operational Overview

The development, release and maintenance of RStudio Professional Products is a collaborative process.

Most communication among RStudio development team members take place electronically via email and similar means. A non-public email list provides a common forum for discussions along with video conferencing and instant messaging tools, such as Slack.

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 Professional Product 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.

3.2 Source Code Management

All of RStudio's commercial source code is managed via git, a free and open source distributed version control system. In particular, RStudio subscribes to GitHub, a cloud-based commercial git provider. The RStudio Professional Product repositories are access controlled, such that only members of RStudio development team have write access to the source code tree.

Source code for the open source versions of RStudio Server, RStudio Desktop, and Shiny Server, which have substantial overlap with the corresponding professional products are available at RStudio's GitHub page at <http://github.com/rstudio>.

Source code workflows for RStudio Professional Products follow standard software engineering processes. The current release branch is only for code that has been thoroughly reviewed and tested. The release branch allows changes only for bug fixes and minor feature enhancements. Major features are subjected to quality assessment and validation in a separate branch before becoming the next release.

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

The release histories and notes for each product are tracked on the product page: <https://docs.rstudio.com/products.html>. Additionally, detailed explanations of new features in each release are shared on the RStudio Blog.

3.3 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 within the source code control system to validate the code at a component level. Additional tests in other source control repositories and manual testing run by the Quality Assurance team prior to release validate the end-to-end functionality of RStudio Professional Products. Any errors noted during this testing are resolved prior to release.

Additional testing of RStudio Professional Products is solicited from the user community during “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 GitHub issues, <https://community.rstudio.com>, and <http://support.rstudio.com> where users report issues and seek support. This process enables a wider array of feature testing and further increases the likelihood of resolving issues prior to the release of a stable version.

3.4 Release Cycle

Once the in-development version of RStudio Professional Products have been approved for release by a designated Release Manager, a public announcement is made via the RStudio blog at <https://blog.rstudio.com/>.

Pre-built executable binary install files for RStudio Professional Products are made available at [RStudio.com](https://www.rstudio.com) for common operating system and CPU architectures. Availability for particular operating systems depends on the product, but RStudio Professional Products exist for Linux, Windows and MacOS platforms.

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

Additional instructions regarding the use of RStudio Professional Products, installation requirements and platform and operating system related issues are extensively documented in the Administration Guides for each product, which are available for each version released at <https://docs.rstudio.com/products.html>.

3.5 Availability of Current and Historical Archive Version

Daily builds of binaries for all RStudio products (professional and open source) are available at <https://dailies.rstudio.com/>.

3.6 Maintenance, Support, and Retirement

Each Released Version of an RStudio Professional 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. The platforms covered under RStudio’s support agreement, along with the schedule for platform end of life and RStudio end of support are displayed on the RStudio website (<https://www.rstudio.com/about/platform-support/>).

Purchasers of RStudio Professional Products have access to a variety of support resources. Aside from access to the professional support detailed in the RStudio Support Agreement (<https://www.rstudio.com/about/support-agreement/>), there are a variety of public support channels available to the users of RStudio Professional Products, including the RStudio Blog (<https://blog.rstudio.com>), public support articles (<https://support.rstudio.com>), and the administration and user guides for each product (<https://docs.rstudio.com/products.html>).

3.7 Qualified Personnel

All members of RStudio’s development team hold qualifying degrees and/or prior development experience, many with Ph.D. and/or Master’s degrees from accredited academic institutions. Many have published in peer reviewed journals and 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
- Southern Methodist University
- Rice University
- Worcester Polytechnic Institute
- University of British Columbia
- Duke University
- University of Edinburgh
- University of Copenhagen
- Technical Univeristy of Denmark

3.8 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.

3.9 Disaster Recovery

RStudio Professional Products are installed on customer systems and are therefore subject to customer disaster recovery practices. RStudio does not operate computational servers that would require continuous uptime.

For delivery of RStudio Professional Product purchases and upgrades, we rely on Amazon for availability of our binaries and disaster recovery practices.

For development of RStudio Professional 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.

As a further precaution and primarily for the benefit of RStudio, we provide source code versions of RStudio Professional Products to a third party source code escrow service.

4 21 CFR Part 11 Compliance Functionality

4.1 Overview

Within the regulated domain, RStudio Professional 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 Professional 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 Professional 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)

4.2 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 Professional Products are not intended to create, maintain, modify, or delete Part 11 relevant *records* but to perform calculations, draw graphics, deliver interactive web experiences, and serve programming dependencies using the R statistical programming language. Where RStudio Professional 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 independent of the use of RStudio Professional Products, 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 Professional Products.

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

4.3 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 Professional Products must be stored in a manner that enables accurate and ready retrieval.

RStudio Professional Products are not intended to create, maintain, modify or delete Part 11 relevant *records* but to perform calculations, draw graphics, deliver interactive web experiences, and serve programming dependencies using the R statistical programming language.

Therefore, *records* created by RStudio Professional 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.

4.4 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 Professional 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.

As a further measure, RStudio's server-based Professional Products (RStudio Server Pro, RStudio Connect, RStudio Package Manager, and Shiny Server Pro) further restrict access to authorized users by enabling or leveraging existing authentication practices including SAML, LDAP, ActiveDirectory, Google Authentication, PAM authentication and sessions, and SSL.

4.5 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 Professional Products are not intended to create, maintain, modify or delete Part 11 relevant records but to perform calculations, draw graphics, deliver interactive web experiences, and serve programming dependencies using the R statistical programming language.

Where RStudio Professional Products' use may be interpreted as creating records, however, its use of the language R can support audit trail creation within the record using built-in functions to 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 Professional 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.

As further measures, RStudio Professional Products provide logging and auditing 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.

4.6 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 Professional 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).

4.7 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 Professional Products.

4.8 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 Professional 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 Professional Products enhance the host environment capabilities as discussed previously, notably in sections 11.10(d) and 11.10(f).

RStudio Professional Products are not intended to create, maintain, modify or delete Part 11 relevant *records* but to perform calculations, draw graphics, deliver interactive web experiences, and serve programming dependencies using the R statistical programming language.

4.9 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. RStudio Professional Products are deployed at the discretion of users, so the creation and dissemination of any such policies are incumbent on user organizations.

RStudio Professional Products are not intended to create records but to perform calculations, draw graphics, deliver interactive web experiences, and serve programming dependencies using the R statistical programming language. Following from this, they are not intended to allow for signature of *records*.

4.10 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 Professional 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 Professional Products' source code.

This documentation is provided to RStudio Professional Products users in electronic formats at <https://docs.rstudio.com/products.html>.

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

4.11 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 Professional 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 Professional Products' user to ensure that the appropriate safeguards are implemented for a particular hosting system.

5 Bibliography

[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.