

Data Science platform with RStudio Server and GitLab

Stricter security policies have increasingly limited the local execution of R and RStudio at Agroscope. As part of an architecture redesign, IT Agroscope has decided to centralise these applications.

With RStudio Server and GitLab, b-data GmbH introduced software that is fully integrated into the architecture. The combination of these applications will enable researchers to introduce a unified handling of data analyses and create reproducible results in the future.

Client: Agroscope – Swiss research into agriculture, nutrition and the environment

Agroscope is the Swiss centre of excellence for agricultural research, and is affiliated with the Federal Office for Agriculture (FOAG). Agroscope makes an important contribution to a sustainable agriculture and food sector as well as to an intact environment, thereby contributing to an improved quality of life.

Agroscope researches along the entire value chain of the agriculture and the food sector. Its goals are a competitive and multifunctional agricultural sector, high-quality food for a healthy diet, and an intact environment. In pursuing these aims, the research institute gears itself to the needs of its service recipients.

Agroscope has outlined 17 strategic research fields for the research and development activities. Topic-related research across the different sites is of crucial importance here.

Requirements of the client: «Work as before» and integration into the IT architecture

The needs of the researchers could be summed up in one statement: «We would like to continue working as before». But what does this mean in detail?

1. Provide the functionality of RStudio Desktop.
2. Assign resources (CPU, RAM) tailored to need.
3. Allow the installation of your own R packages.
4. Have several R versions to choose from and
5. allow access to the existing data storage.

The basic requirement of IT was the full integration into its architecture, i.e.

1. Virtualization on VMWare vSphere
2. Integration into the Agroscope domain
3. Interoperability with Active Directory

To verify all the requirements of research and IT, the project has been split up into three phases:

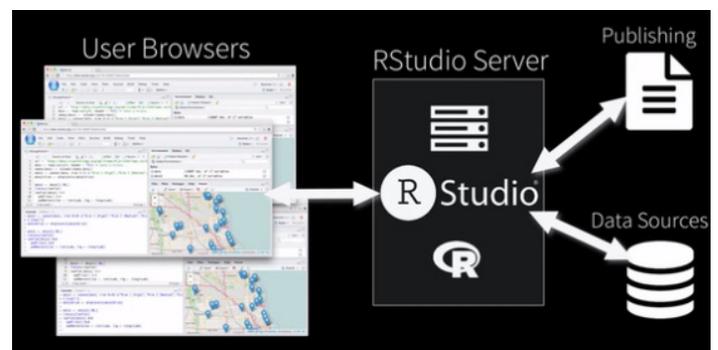
1. **Test phase:** Testing of the functionalities by a core team of researchers.
2. **Pilot phase:** Checking IT security policies and integration with internal services.
3. **Integration:** Acceptance by IT and subsequent release for productive operation.

Fundamental part of this project was a comprehensive documentation of the server configurations as well as detailed training materials for the introduced applications.

Solution from b-data: RStudio Server Pro and GitLab CE

Since most researchers are already familiar with RStudio Desktop, the decision in favour of RStudio Server Pro was

made quickly. In combination with GitLab CE, it will also be possible to ensure the reproducibility of research projects. In order to minimize the effort for IT in rolling out the pilot environment as well as productive servers, the virtual machines were made available in the Open Virtual Machine Format (OVF).



Conclusion: Future-proof applications as basis for Reproducible Research

Ubuntu, RStudio Server and GitLab provide Agroscope with well-established and state-of-the-art applications. With the introduction of RStudio v1.2, researchers will also be free to perform their analyses in R, Python or a combination of both languages.

The demands on scientific data analysis and its documentation are constantly increasing. With the technologies implemented, researchers have up-to-date tools to apply *Reproducible Research*:

1. Versioning of the code (RStudio Project + Git)
2. Sharing the analyses with collaborators (GitLab)
3. Central data storage for the raw data

«Olivier understood the complex needs of the researchers and was able to fully consider them in the integration and operational concept. Of course, the migration brought about changes, but overall this resulted in new opportunities for scientific data analysis.»

– Manuel K. Schneider, Research Group Forage Production and Grassland