

# Data Science Plattform mit RStudio Server und GitLab

**Verschärfte Sicherheitsrichtlinien haben dazu geführt, dass die lokale Anwendung von R und RStudio bei Agroscope zunehmend eingeschränkt wurde. Im Rahmen einer Architekturanpassung hat sich die IT Agroscope dazu entschlossen, diese Fachapplikationen zu zentralisieren.**

**Mit RStudio Server und GitLab führte die b-data GmbH Anwendungen ein, welche vollständig in die Architektur integriert sind. Die Verknüpfung dieser Applikationen ermöglicht den Forschenden zukünftig die Einführung einer einheitlichen Handhabung von Datenanalysen sowie die Erzeugung reproduzierbarer Resultate.**

## Auftraggeber: Agroscope – Schweizer Forschung für Landwirtschaft, Ernährung und Umwelt

Agroscope ist das Kompetenzzentrum des Bundes für landwirtschaftliche Forschung und ist dem Bundesamt für Landwirtschaft (BLW) angegliedert. Agroscope leistet einen bedeutenden Beitrag für eine nachhaltige Land- und Ernährungswirtschaft sowie eine intakte Umwelt und trägt damit zur Verbesserung der Lebensqualität bei.

Die Forschung erfolgt entlang der gesamten Wertschöpfungskette der Land- und Ernährungswirtschaft. Ziele sind eine wettbewerbsfähige und multifunktionale Landwirtschaft, hochwertige Lebensmittel für eine gesunde Ernährung sowie eine intakte Umwelt. Dabei richtet sich die Forschungsanstalt auf die Bedürfnisse ihrer Leistungsempfänger aus.

## Anforderungen des Kunden: «Arbeiten wie bisher» und Einbindung in die IT Architektur

Die Bedürfnisse der Forschenden liessen sich in einer Aussage zusammenfassen: «Wir möchten so weiterarbeiten können, wie bisher». Doch was bedeutet dies im Detail?

1. Den Funktionsumfang von RStudio Desktop zur Verfügung stellen.
2. Die Ressourcen (CPU, RAM) bedarfsgerecht zuweisen.
3. Die Installation eigener R Pakete erlauben.
4. Mehrere R Versionen zur Auswahl haben und
5. den Zugriff auf die bestehende Datenablage ermöglichen.

Die grundlegende Anforderung der IT war die vollständige Einbindung in ihre Architektur, d.h.

1. Virtualisierung auf VMware vSphere
2. Einbindung in die Agroscope-Domäne
3. Kompatibilität mit Active Directory

Um sämtliche Anforderungen von Seiten Forschung und IT zu verifizieren, wurde das Projekt in drei Phasen unterteilt:

1. **Testphase:** Testen der Funktionalitäten durch ein Kernteam von Forschenden.
2. **Pilotphase:** Überprüfen der IT-Sicherheitsrichtlinien und Anbindung an die internen Services.
3. **Integration:** Abnahme durch die IT und anschliessende Freigabe für den produktiven Betrieb.

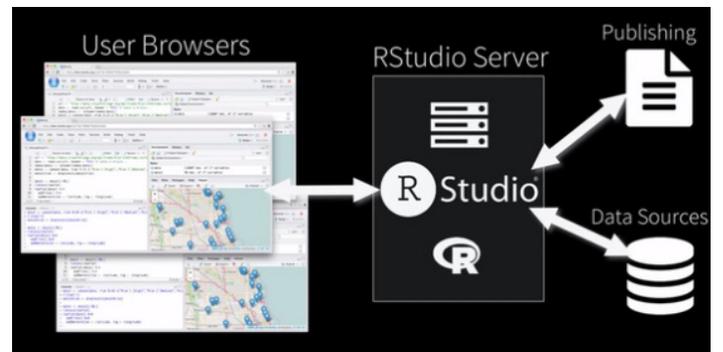
Wichtiger Bestandteil dieses Projektes waren eine umfassende Dokumentation der Serverkonfigurationen sowie ausführliche Schulungsunterlagen zu den eingeführten Anwendungen.

## Lösung von b-data: RStudio Server Pro und GitLab CE

Da die meisten Forschenden bereits mit RStudio Desktop vertraut sind, war der Entscheid zu Gunsten von RStudio

Server Pro rasch gefällt. In Kombination mit GitLab CE wird zudem die Möglichkeit geboten, die Reproduzierbarkeit von Forschungsprojekten sicherzustellen.

Um den Aufwand für die Inbetriebnahme der Pilotumgebung sowie produktiven Server für die IT minimal zu halten, wurden die virtuellen Maschinen im Open Virtual Machine Format (OVF) zur Verfügung gestellt.



## Fazit: Zukunftssichere Anwendungen als Grundlage für Reproduzierbare Forschung

Mit Ubuntu, RStudio Server und GitLab stehen Agroscope etablierte Anwendungen zur Verfügung, die stets auf dem aktuellen Stand der Technik gehalten werden. Mit der Einführung von RStudio v1.2 wird es den Forschenden ferner freistehen, ihre Analysen in R, Python oder in Kombination beider Sprachen durchzuführen.

Die Anforderungen an die wissenschaftliche Datenanalyse und ihre Dokumentation steigen stetig. Mit den implementierten Technologien haben die Forschenden zeitgemässe Werkzeuge, um *Reproduzierbare Forschung* zu betreiben:

1. Versionierung des Codes (RStudio Projekt + Git)
2. Teilen der Analysen mit Mitarbeitern (GitLab)
3. Zentrale Datenablage für die Rohdaten

«Olivier verstand die vielschichtigen Bedürfnisse der Forschenden und konnte diese vollständig im Integrations- und Betriebskonzept berücksichtigen. Natürlich brachte die Migration Umstellungen mit sich, aber insgesamt ergaben sich daraus neue Möglichkeiten für die wissenschaftliche Datenanalyse.»

– Manuel K. Schneider, Forschungsgruppe Futterbau und Graslandssysteme