

HERZLICH WILLKOMMEN

# [R]EADY FOR PRODUCTION DATA-SCIENCE-EVENT 2019



eoda  
daten ► wissen ► nutzen



# Über eoda | DIE DATA SCIENCE SPEZIALISTEN



## FACTS

- Gegründet 2010 | Sitz in Kassel
- Interdisziplinäres Team
- 50 Mitarbeiter



## MISSION

- Entwicklung von Analyse- und Prognosemodellen
- Evidenzbasierte Management Summaries
- Data-Science-Enablement
- Operationalisierung von Data Science



# UNSER PORTFOLIO

Beratung & Projekte

**Use-Case-Beratung**

**Projektrealisierung**

**Coaching & Support**

**aicon | Assessment**

**aicon | Implementierung**

**aicon | Betrieb & Support**

Technologie & Infrastruktur



**eoda**  
daten ► wissen ► nutzen

Wissenstransfer

**R-Training**

**Python-Training**

**Impulsvorträge**



Software

**DATA-SCIENCE-EVENT 2019**

# INFRASTRUKTUR: eoda | aicon

analytic infrastructure consulting

Best-Practice-Ansatz für den professionellen Einsatz von Data Science

## eoda | aicon Betrieb

- Überwachung, Wartung und Support

**Betrieb**

## eoda | aicon Implementierung

- Einrichtung des Analyse-Environments

**Implementierung**

**Realisierungskonzept**

**Workshop**

**Bestandsaufnahme**

## eoda | aicon Assessment

- Überblick über den Status quo Ihrer Analyselandschaft
- Individuelle Handlungsempfehlungen



# INFORMATIONEN ZUR LOCATION

- Lunch im Nebenraum
- BarCamp findet sowohl im Historischen Festsaal, als auch im Nebenraum statt
- Bitte die Fenster geschlossen halten
- Eingangstüren schließen automatisch
- WLAN 2,4G: SE-Baustelle
- WLAN 5G: SE-Baustelle-5G
- PASSWORT: mD&c9KVA!MF

DATA-SCIENCE-EVENT 2019



# DIE AGENDA

10:00-10:20

## BEGRÜßUNG UND KURZVORSTELLUNG

Oliver Bracht | Chief Data Scientist | eoda GmbH

Dave Hurst | Director Business Development | RStudio

10:20-10:50

## R IN PRODUKTIVUMGEBUNGEN

Das richtige Zusammenspiel zwischen Data Science und IT

Bastian Wießner | Solutions Engineer | eoda GmbH

10:50-11:05 | KAFFEE & SNACKS

11:05-12:00

## SUCCESS STORY: AUFBAU EINER ANALYSEUMGEBUNG BEI DER REWE INTERNATIONAL AG

Alex Burisch | IT Organisator | REWE International AG

Bastian Wießner | Solutions Engineer | eoda GmbH

DATA-SCIENCE-EVENT 2019

12:00-13:00 | LUNCH

13:00-14:00

## RSTUDIO TALK

Deploying Models in Production using REST APIs

Andrie de Vries | Solutions Engineer | RStudio

14:00-14:30 | KAFFEE & SNACKS

14:30-16:30

## BARCAMP

IM ANSCHLUSS | VERABSCHIEDUNG



eoda  
daten ► wissen ► nutzen



# DAS BARCAMP

	TRACK 1	TRACK 2	TRACK 3
14:30-15:30	Entwicklung von Shiny Apps	Multiple Languages (Tutorial)	Package Management
15:30-16:30	Versionsverwaltung in Data Science	Plumber APIs in R (Tutorial)	



# R in Produktivumgebungen

-

Das richtige Zusammenspiel zwischen  
Data Science und IT

Bastian Wießner | Solutions Engineer | eoda GmbH

DATA-SCIENCE-EVENT 2019





# AGENDA

Eine Frage  
der Mentalität



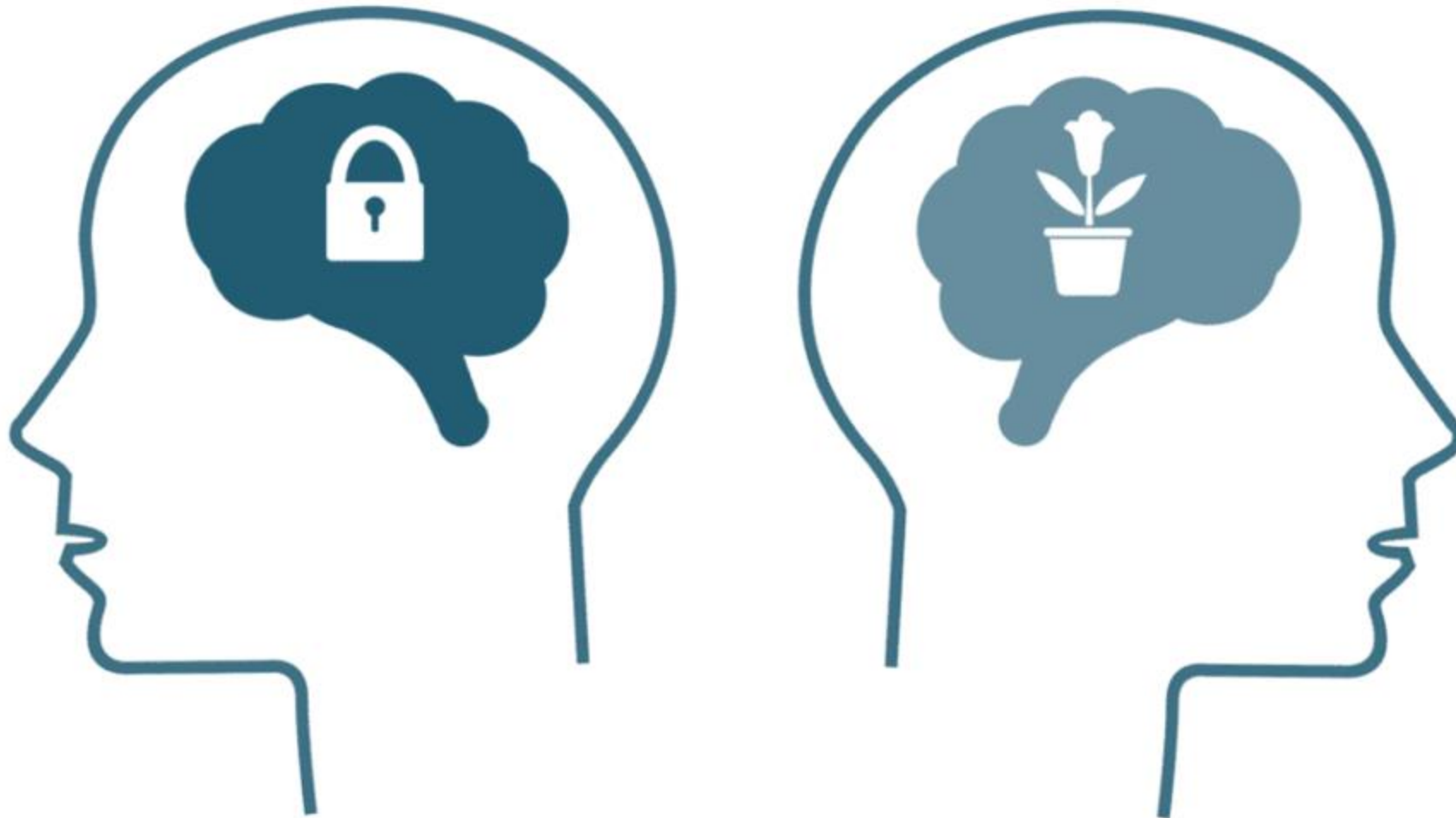
„Produktiv“ gehen



produktivR



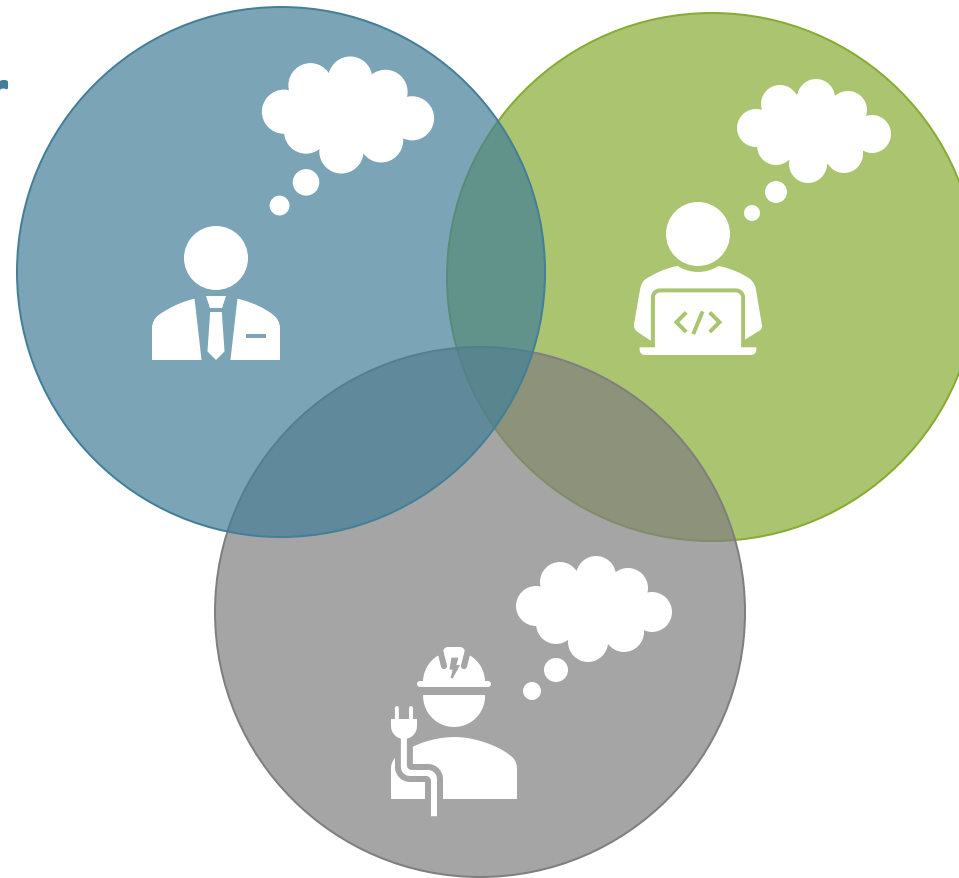
# EINE FRAGE DER MENTALITÄT



# DIE IDENTIFIKATION DER GEMEINSAMEN SCHNITTMENGE

**Business User**

**Data Science**



**IT Betrieb**



# IT VS. DATA SCIENCE

## IT

- Sicherheit
- Zuverlässigkeit
- Wartbarkeit





# IT VS. DATA SCIENCE

## Data Science

- Agilität
- Wunsch nach Bleeding Edge Technologie
- Gestaltungsfreiheit



# IT VS. DATA SCIENCE

## Data Science

- Agilität
- Wunsch nach Bleeding Edge Technologie
- Gestaltungsfreiheit

Data  
Science

≠

Software  
Engineering





# BUSINESS VS. IT

## Business

- Großes Vertrauen in die Ergebnisse aus Data Science
- Keine / geringe Ausfallzeiten

## IT

- Wenig Vertrauen, da u.a.
  - Kein Test-Team
  - Kein Requirements Engineering
  - Unübersichtliche Abhängigkeiten
- Governance Barriere
  - Public CRAN Server mit ~14.500 Paketen



# PERSPEKTIVEN AUF ÄNDERUNGEN



Data Scientist

Chancen  
Agilität  
Freiheit



Risiko  
Stabilität  
Regeln



IT-Operator



# AGENDA

Eine Frage  
der Mentalität



„Produktiv“ gehen



produktivR



# WANN IST EIN PROZESS „PRODUKTIV“?

## Merkmale produktiver Analytik-Prozesse:

- Zyklische Analysen, die zeitlich gesteuert sind
- oder Analysen, die stets weiterentwickelt werden
- und Mehrwerte für das Unternehmen bringen





# CHECKLISTE FÜR PRODUKTIVE UMGEBUNGEN

- ✓ IT-Wissen
- ✓ Staging Server
- ✓ Blue / Green deployment
- ✓ Integration Environment
- ✓ Qualitätssicherheit Teams
- ✓ Automatisiertes Testing
- ✓ CI / CD Pipelines
- ✓ Versionskontrolle
- ✓ Changemanagement
- ✓ Netzwerkmanagement
- ✓ Securitymanagement
- ✓ DevOps
- ✓ 24-7 support
- ✓ Releasemanagement
- ✓ Containerisierung

# CHECKLISTE FÜR PRODUKTIVE UMGEBUNGEN

✓ IT-Wissen

✓ Changemanagement

**Sollte das jeder Data Scientist können müssen?**

**NEIN!**

✓ Blue / Green deployment

✓ Securitymanagement

✓ Integration Environment

✓ DevOps

✓ Preprod Environment

 **Es geht darum Brücken zu bauen, um einen**

**Dialog zwischen IT und Data Science zu ermöglichen**

✓ Automatisiertes Testing

✓ Containerisierung

✓ CI / CD Pipelines



# CHECKLISTE FÜR R IN PRODUCTION

- ✓ Testing Strategie
- ✓ Deployment Prozess
- ✓ Release Prozess
- ✓ Change Management
- ✓ Security review
- ✓ Versionskontrolle
- ✓ Coding Guideline
- ✓ Support
- ✓ R-Version
- ✓ Validierte Pakete
- ✓ System Abhängigkeiten
- ✓ Logging

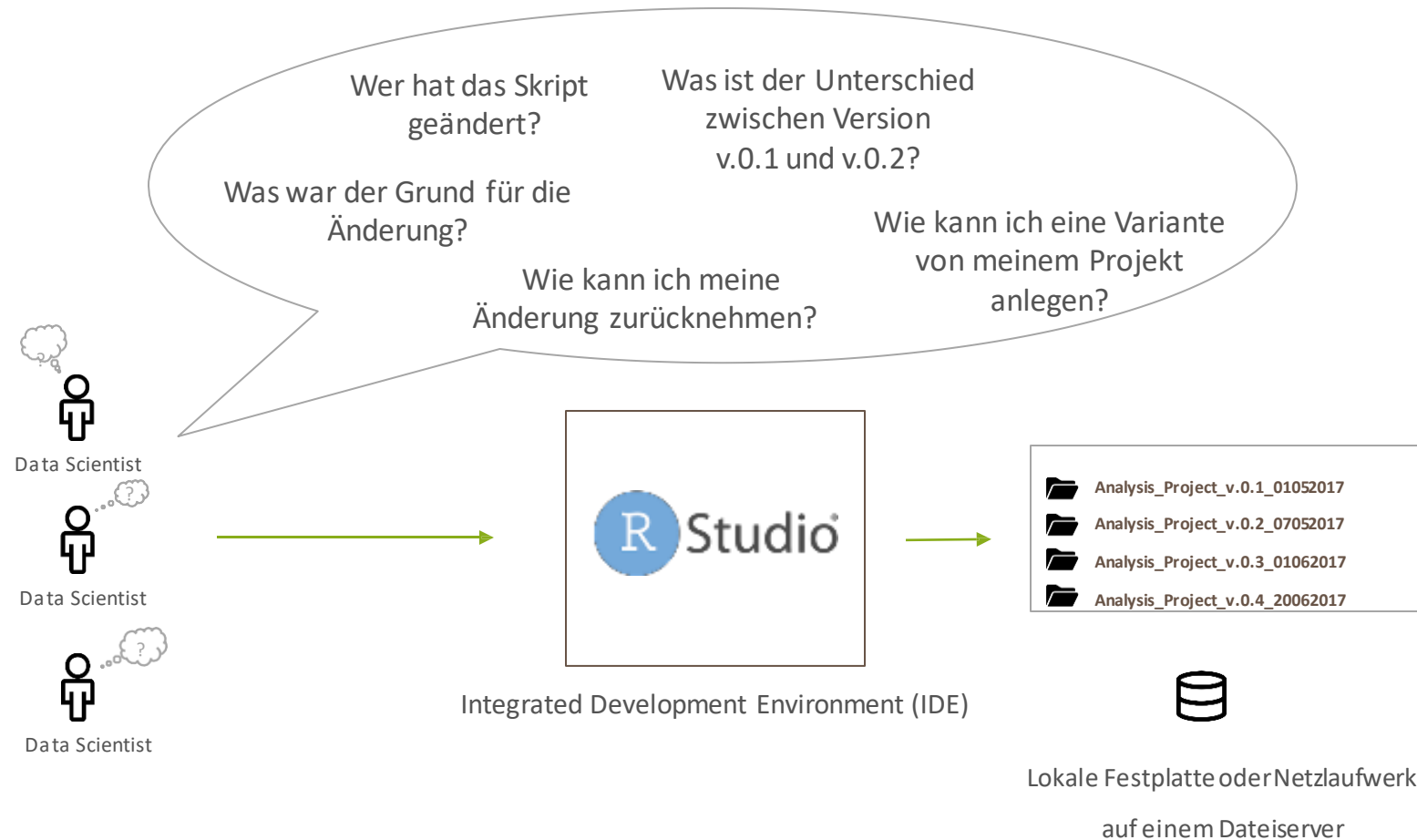


# WIE SIEHT DER TYPISCHE R-EINSATZ IN UNTERNEHMEN AUS?

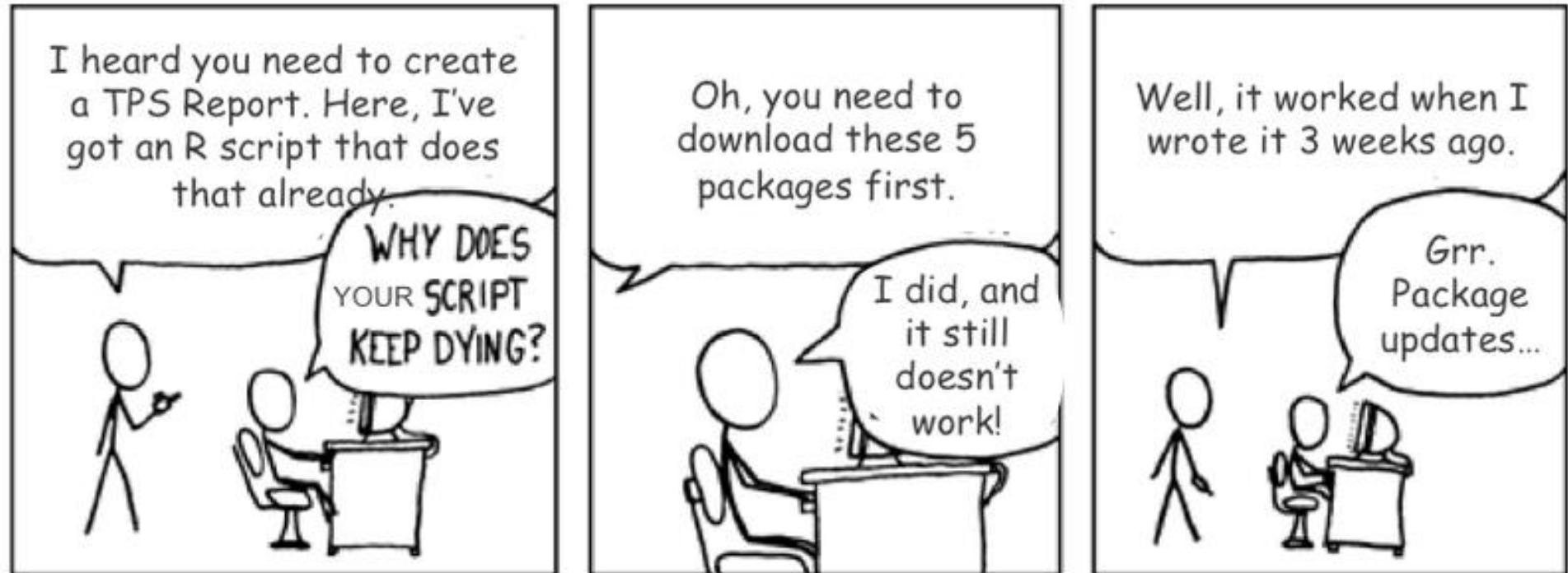
- „Lab Teams“ vs. „Data-Science-Teams“
- Dezentralisierte Clients
  - Unterschiedliche Betriebssysteme
  - Verschiedene Versionen von Komponenten
  - Unterschiedliche Bibliotheken
  - Teilweise unversioniertes Arbeiten
  - Verschieben von Dokumenten /Skripten via Mail
- Keine Qualitätssicherung
- Keine CI/CD Pipelines



# PROBLEM EINS: KEINE VERSIONSVERWALTUNG

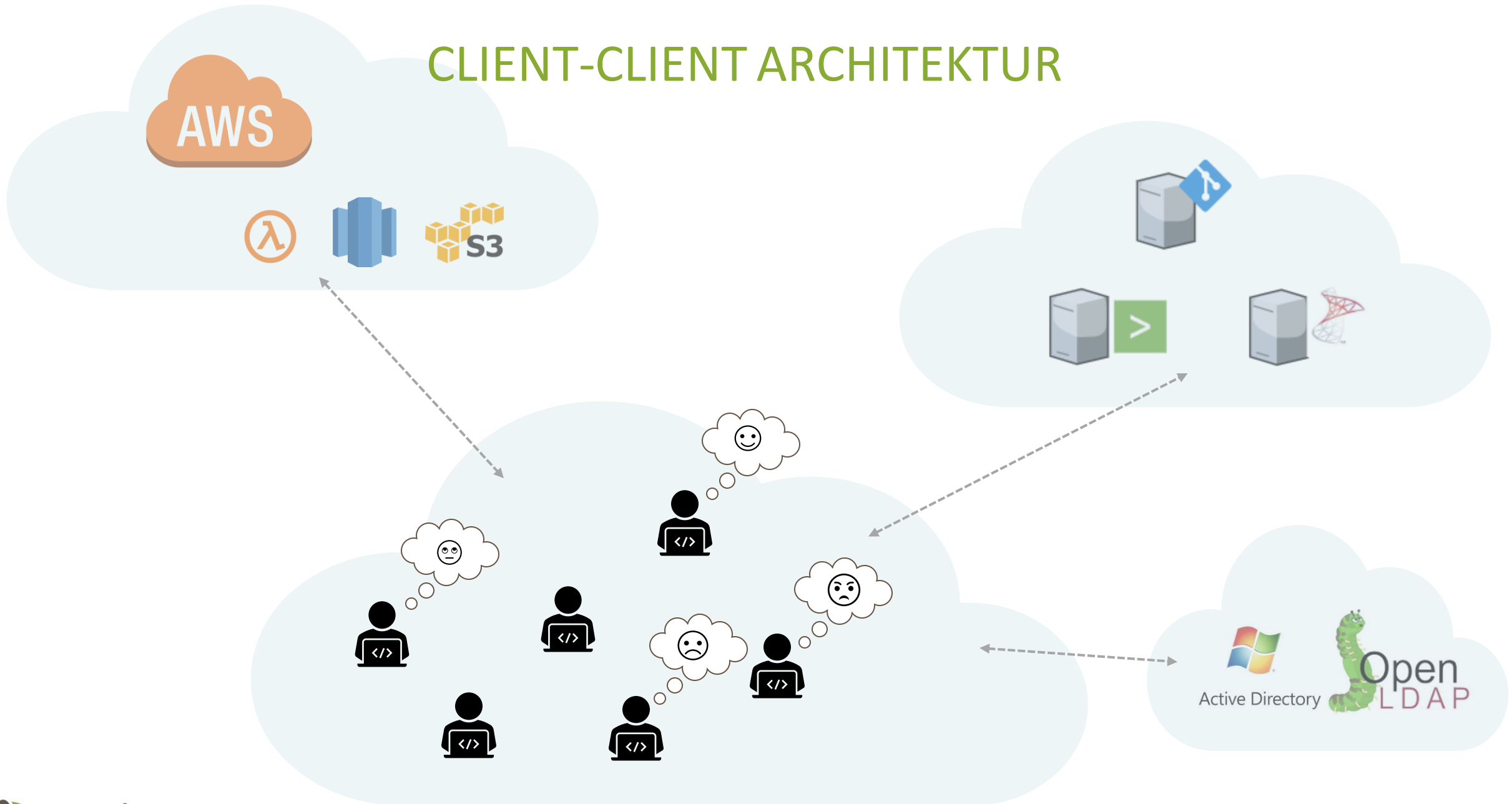


# PROBLEM ZWEI: UNTERSCHIEDLICHE PAKETVERSIONEN





# CLIENT-CLIENT ARCHITEKTUR



# AGENDA

Eine Frage  
der Mentalität



„Produktiv“ gehen



**produktivR**



# WIE FUNKTIONIERT R IM PRODUKTIVEINSATZ?

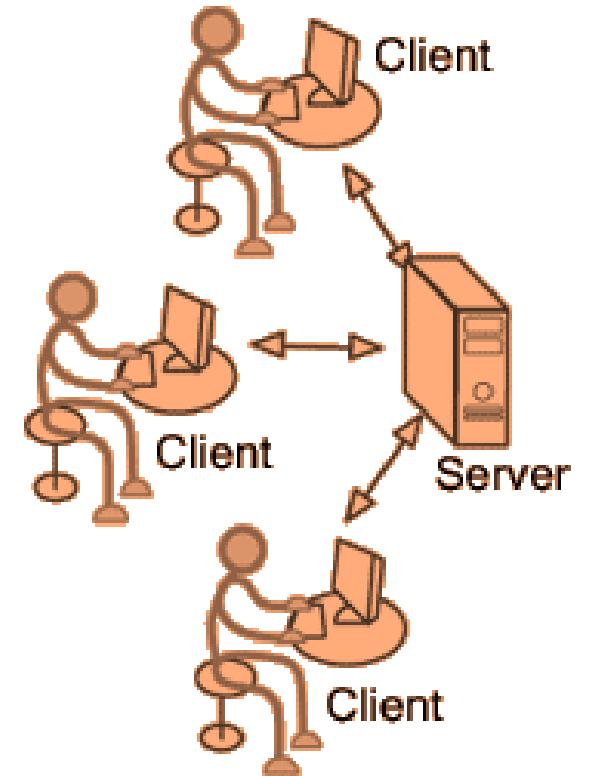
- R ist eine leichtgewichtige Skriptsprache
  - Leicht zu portieren / schmaler Festplattenspeicher nötig
  - Hat kaum Abhängigkeiten zum Betriebssystem
  - Leicht zugänglich für erste Erfolge
- Sehr gutes Paketökosystem
- Einfach zu bedienendes Unit Test Framework
- Schneller Entwurf von (shiny) Dashboards



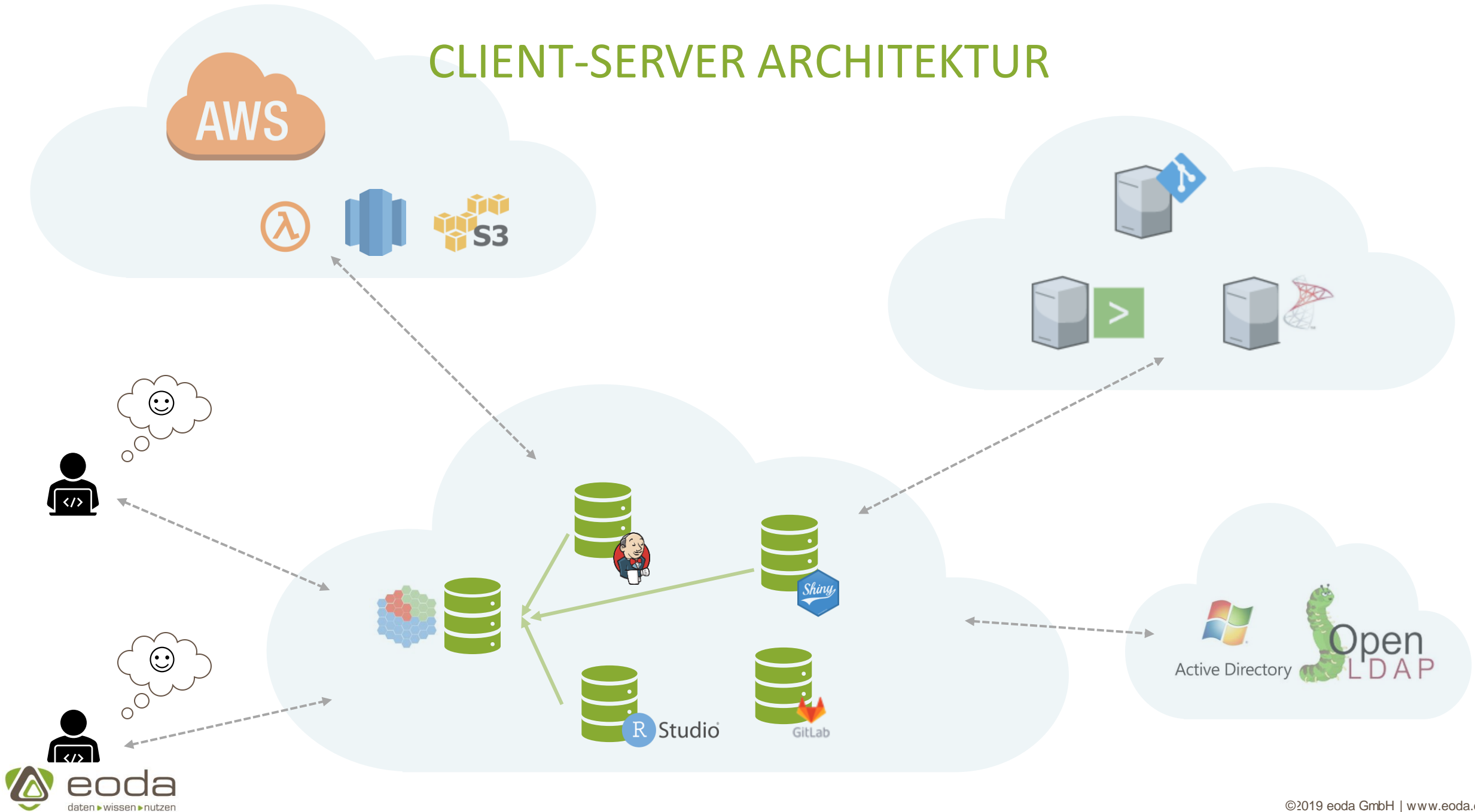


# WARUM EINE CLIENT-SERVER ARCHITEKTUR?

- Einmalige serverseitige Lösung
- Standardisierte Toolsets
- Zentrale Wartung der Infrastruktur + installierte Komponenten
- Vereinfachter Austausch von Daten zwischen den Kollegen
- Bessere Performance gegenüber Desktop-Rechnern



# CLIENT-SERVER ARCHITEKTUR



# Enablement der Akteure

- Enablement der IT
- Enablement der R-Administratoren
- Enablement der Data-Science-Abteilung





## Field Guide to the R Ecosystem

<https://fg2re.sellorm.com/>





*Die Data Science Spezialisten.*

**eoda GmbH**  
Universitätsplatz 12  
34127 Kassel  
[www.eoda.de](http://www.eoda.de)  
[sales@eoda.de](mailto:sales@eoda.de)  
+49 561 202724-51



@eodaGmbH



[blog.eoda.de](http://blog.eoda.de)



@eodaGmbH



eoda GmbH