Beyond prototyping: Best practice for R in critical enterprise environments
Types of changes

- Operational System
- Interpreter
- Packages and their dependencies
- Scripts
- Structure of the Data
Components

- Version Control + Branching Strategy
- Staging Architecture
- Continuous Integration for automated tests
- Package Management
Branching Strategy based on GitFlow

- **main branches**
  - **master**
  - **develop**

- **supporting branches**
  - **release/1.0.0**
  - **hotfix/bug-433**
  - **release/1.1.0**

- **New Release**
  - Tag: 1.0.0
  - Tag: 1.0.1
  - Tag: 1.1.0

- **Bugfixing**
  - New Release
  - Tag: 1.1.0

- **User Acceptance Test**
  - Release
  - Tag: 1.0.0

- **time**

© 2017 eoda GmbH
Continuous Integration

Data Science Team

- Check-in / push to remote
- Managed remote Repository
- Continuous Integration Server
- Webhook trigger
- Start CI pipeline
- local branches

Development
- develop
- test
- production

Test
- test
- hotfix
- master

Feedback
- Feedback to Repository about Success
- Put new Version on respective System, if all steps before succeed

Prepare
- Starting R Session

Test
- Perform Unit-, Integration-, Functional- and Performance-Tests

Build Package (if applicable)

Checkout
- Clean working directory
- Copy Branch to working directory

Deploy
- Build Package (if applicable)
- Feedback to Repository about Success
- Put new Version on respective System, if all steps before succeed
Standard Paketmanagement

Pakete:
- dplyr 0.7
- ggplot 2.2
- Hmisc 4.0
- caret 6.0

Pakete:
- dplyr 0.6
- ggplot 2.2
- caret 5.8

Pakete:
- dplyr 0.7
- ggplot 1.4
- Hmisc 1.1

Pakete:
- dplyr 0.3
- ggplot 1.8
- Hmisc 6.0
- caret 2.1

Pakete:
- dplyr 0.8
- ggplot 1.4
- nnet 7.0

CRAN

User 1

User 2

User 3

Projekt 1
Script 1
Script 2
Script 3

Projekt 2
Script 1
Script 2

Projekt 3
Script 1
Script 2
Script 3
Script 4
Decentralized Package Management (packrat)

Production
Test
Development

CRAN

User 1

R
Pakete:
3.3

User 2

R
Pakete:
3.1

User 3

R
Pakete:
3.2

Pakete:
2.8

Projekt 1

dplyr 0.7
ggplot 2.2
Hmisc 4.0
caret 6.0

Script 1
Script 2
Script 3

Projekt 2
dplyr 0.6
ggplot 2.2
caret 5.8

Script 1
Script 2

Projekt 3
dplyr 0.7
ggplot 1.4
Hmisc 1.1

Script 1
Script 2
Script 3
Script 4
Centralized Package Management (miniCran)

**Pakete:**
- **Pakete:** ggplot 2.2
caret 6.0
car 2.1
- **Pakete:** ggplot 2.2
caret 6.0
car 2.1
- **Pakete:** ggplot 2.2
caret 6.0
car 2.1

**Projekt 1**
- Script 1
- Script 2
- Script 3

**Projekt 2**
- Script 1
- Script 2

**Projekt 3**
- Script 1
- Script 2
- Script 3
- Script 4

* Local Repository
Decentralized Package Management

Updates

If there is an critical update in one of the packages, this update needs to be performed for each Projekt. Who takes care about it?
Centralized Package Management

Updates

The Data Science Admin performs an update on a dedicated test system first and checks whether or not the projects will cause issues with the new package version. If everything ok, The update will go the „regular“ LRAN and will be spread out automatically.
Meet us on the third floor

Die Data Science Spezialisten.

eoda GmbH
Universitätsplatz 12
34127 Kassel
www.eoda.de
info@eoda.de
+49 561 202724-40

@eodaGmbH
blog.eoda.de
@eodaGmbH
eodaGmbH