Bernhard Meindl
Statistics Austria (QM)

Cham/Zug
November 2019

Managing a modern R-Environment

What's the plan?

- Sharing the experiences
  - Switching from Rstudio Desktop $\longrightarrow$ Rstudio Server
  - Improving reproducibility
- Discussing our infrastructure (and why we are happy with it)
- Talk about future plans

Starting Point

- Situation until Q1/2018
  - Methods unit prepared **R** installation package
    - R + Packages + Rstudio (Desktop)
    - LaTex + dev-tools (compilers, git, ...)
  - Rollout by IT for all (~100) **R** installations
- Problem
  - difficult to maintain
  - only a single version of R provided at a time
- But
  - Methods unit always had access to a instance running Rstudio-Server
  - very good experiences!

$\longrightarrow$ Idea: let everybody have access to the server version of Rstudio

Timeline (2017)

- Preparation
  - Autumn: internal decision to migrate all **R**-users to Rstudio Server
  - Cooperation with IT-department (important):
    - Cast of a deployment strategy and specific tasks
    - Setup of two (identical) Linux servers (test and production)
    - Setup of Rstudio-Server Pro (with evaluation license)
- Evaluation and Testing
  - (very positive) evaluation of Rstudio-Server
  - Useful features include:
    - per user/group resource allocation
    - multiple **R** versions
    - multiple concurrent sessions, ...

Timeline (2018)

STATISTICS AUSTRIA
The Information Manager

`Training and Adjustments`

- Design of a training course for the new setup
- Rewrite of existing **R**-courses to refer to the new infrastructure

`Start of Migration`

- Transfer (most) users from Desktop to Server
- providing internal **R** packages to users

`But`

- Still maintaining the latest **R** desktop installation as "fallback"
- Identifying use-cases that block removal of old installatiions

`Collaborative Working`

- Testing and Evaulation of Rstudio-Connect

Timeline (2019)

STATISTICS AUSTRIA
The Information Manager

`Implementing Rstudio-Connect`

- Providing examples how to use RSConnect for APIs, reports and shiny apps
- Teaching this to users

`Improving user-experience`

- providing more internal **R** packages
- enhance documentation in confluence (wiki)
- formalize processes (identity management, authentication)
- making use of other available tools (bitbucket, jira, jenkins)

`Hopefully`

- Removing the last Rstudio desktop installations

Setup (1)

STATISTICS AUSTRIA
The Information Manager

Hardware Setup

- 2 virtualized Servers (Test/Production)
  - running Ubuntu 16.04 LTS, Xeon 8 Cores and 128GB RAM
  - easy to scale because it is virtualized
- Problem: estimate peak-usage?
  - estimate 4GB per user and monitor usage/load
  - upgrade users on demand

Software

- Rstudio-Server, Rstudio-Connect, Latex and compilers / devtools
- Additionally: internal Bitbucket installation (git)

Issues, Problems and Feedback

- Using jira to track issues (however, mail and phone popular)

Setup (2)

STATISTICS AUSTRIA
The Information Manager

Why two servers?

- Testserver:
  - mostly used by staff of methods unit
  - we can test new versions of **R** (packages), **Rstudio-Server** and **Rstudio-Connect** before deploying on production server
  - always has latest R-version and devel-version available

Production:

- used by all other users
- latest **R**-version usually is one minor-revision behind the Testserver

Process for User-Setup

- IT creates (system) user on production server
- Authorization to use internal Bitbucket/Jira is granted via identity management tools

STATISTICS AUSTRIA
The Information Manager

Where are we now?

- several training (migration) sessions held
- ~175 **R** users migrated to the server
- overall low number of problems and issues

Current work

- Fine-Tuning: helping users to get accustomed
- Rewriting training materials making use of the new infrastructure
- Identifying use-cases that "block" removal of RStudio-Desktop instances
- Maintaining and updating internal **R** packages

Custom R Packages (1)

STATISTICS AUSTRIA
The Information Manager

`Setup`

- internal artifactory server
- all packages

`Reason`

- One major problem was to give users access to (windows-based) file shares
  - Problem: no mapping between Windows- and Linux user credentials
  - Solution: package `mountSTAT`

`mountSTAT`

- allows to mount windows-based shares with "real" credentials
- secure (encrypted) way of saving credentials using package `secret`
- very easy usage

Custom R Packages (2)

STATISTICS AUSTRIA
The Information Manager

`rinstSTAT`

- allows to install a new **R** version
- allows to upgrade packages for specific **R** versions
- provides a custom API for querying the **R** installations

```
rinstSTAT::install_new_Rversion(
  rvers = "3.6.1", old_library = "/opt/R/3.6.0")
```

`apiSTAT`

- provides API-access to local bitbucket, jira and jenkins installations
- secure credentials using `authSTAT`
- Information about **R** installation

```
apiSTAT::stat_r_api(api = "rinfo", query = "rvers=3.6.1")
apiSTAT::stat_r_api(api = "users")
```

Custom R Packages (3)

`dataSTAT`

- easy access to DB2 databases using odbc and jdbc
  - contains drivers, hardcoded database names
- downloading files from/to sftp-servers
- read data from mainframe and from ms-access using `mdbtools`

`useSTAT`

- provides templates for project codes (folder structure, . . .)
- allows to link projects to git and initializes those projects

Custom R Packages (4)

STATISTICS AUSTRIA
The Information Manager

`sampSTAT`

- creating, modifying, exporting internal sampling frames

`graphSTAT`

- interactively creating standardized graphs

`mzSTAT`

- methods to work with microcensus data

`slideSTAT`

- standardized slides for presentations (like this one)

Future tasks

`Authentication`

- try to come up with a single-sign-on process

`Simplify usage of build-tools`

- help users to make use of jenkins build-server for projects
  - $\longrightarrow$ `buildSTAT`

Experiences and Lessons learned

STATISTICS AUSTRIA
The Information Manager

`Take your time`

- do not rush
- make realistic assumptions on time required for the preparation, implementation and transition phase

`Cooperation with IT`

- at some point you will need their help - make them your friends!

`Friendly users`

- Identify some collegues that are interested in trying out the new environment
- they will provide useful inputs and "free testing"

Experiences (2)

`Be aware of roadblocks`

- Problems you might never have thought of will occur
- People can (and do) crazy stuff with **R**

`Slow down`

- do not try to migrate all users at once - do it in "batches"

`Training and Feedback`

- allow (and reserve) extra time to prepare trainings
- allow time to provide feedback to collegues asking questions

Finish

- I am happy to answer any questions

STATISTICS AUSTRIA
The Information Manager

Please address queries to:
Bernhard Meindl

Contact information:
Guglgasse 13, 1110 Vienna
bernhard.meindl@statistik.gv.at

Managing a modern R-Environment