

Breeding API (BrAPI)

Session C: Data Sharing & Use

Maikel P.H. Verouden

December 12, 2018



Plant Breeding

- Crucial to feeding a rapidly growing population
- Necessity to breed new varieties
 - high productivity with reduced inputs
 - Adapted to new eco-agricultural environments resulting from climate change
- Complex undertaking that necessarily integrates many interrelated disciplines
 - Each their own conventions for data structures and storage (**un-FAIR!**),
 - Increasingly large, multi-faceted datasets

What is



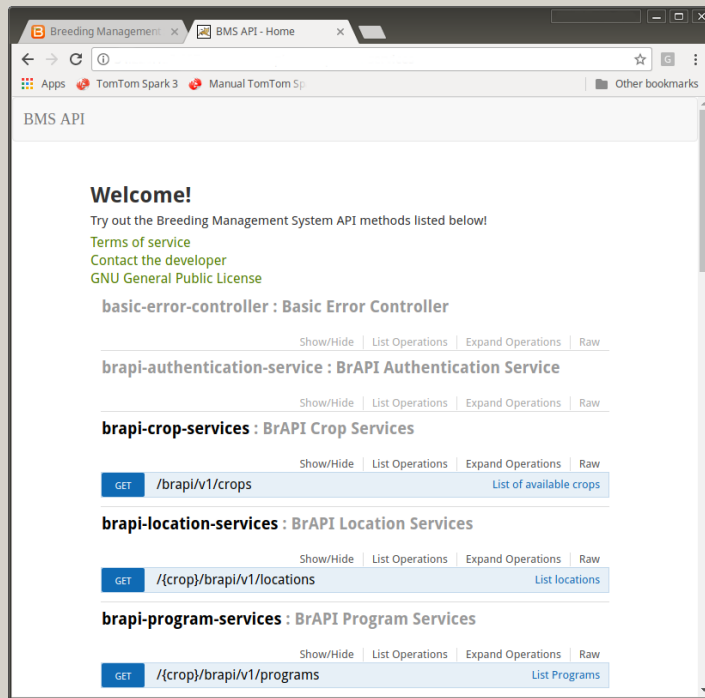
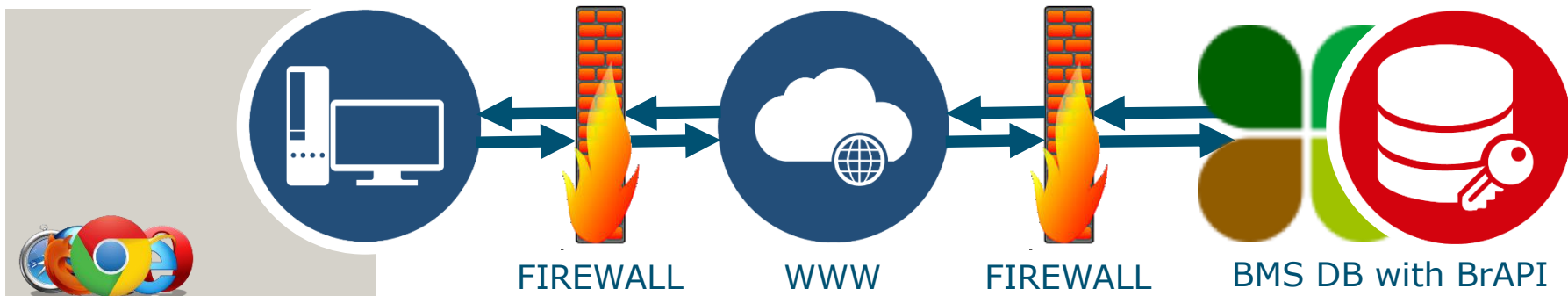
- Breeding Application Programming Interface is an effort to enable **Interoperability** and **Reusability** among plant breeding data-bases.
 - API: a set of subroutine definitions, protocols for building application software.
 - Allow requesting systems to access and manipulate resources.
 - Implemented at the database side and in dedicated application software

Advantages of using



- Specifies a standard interface for plant phenotype /genotype databases to serve their data to crop breeding applications.
 - It is a shared, web-based, open standard
 - Compatibility: MCPD, MIAPPE
 - Optimized for Speed
 - Modular
 - Community-Driven (see brapi.org/partners.php)

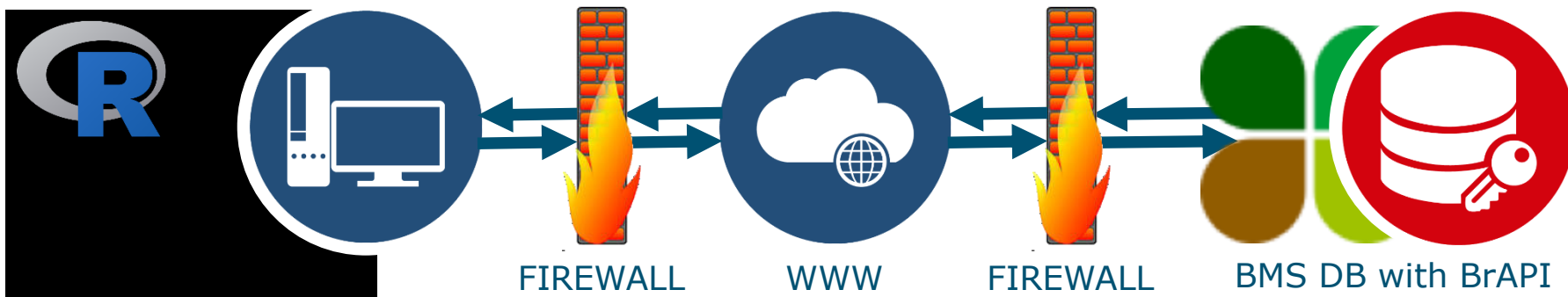
Example Database (BMS) with BrAPI



BMS implemented and validated BrAPI calls:

- Authentication
- List Crops in use
- List Programs
- List Trial summaries
 - Trial can contain multiple studies, i.e. fields
- Get Study/Field phenotypical data as table

Example Software BrAPI R Package (I)



R version 3.5.1 (2018-07-02) -- "Feather Spray"

Copyright (C) 2018 The R Foundation for Statistical Computing

```
> # Load the brapi package into R
```

```
> library(brapi)
```

This is the development version of the 'brapi' package!

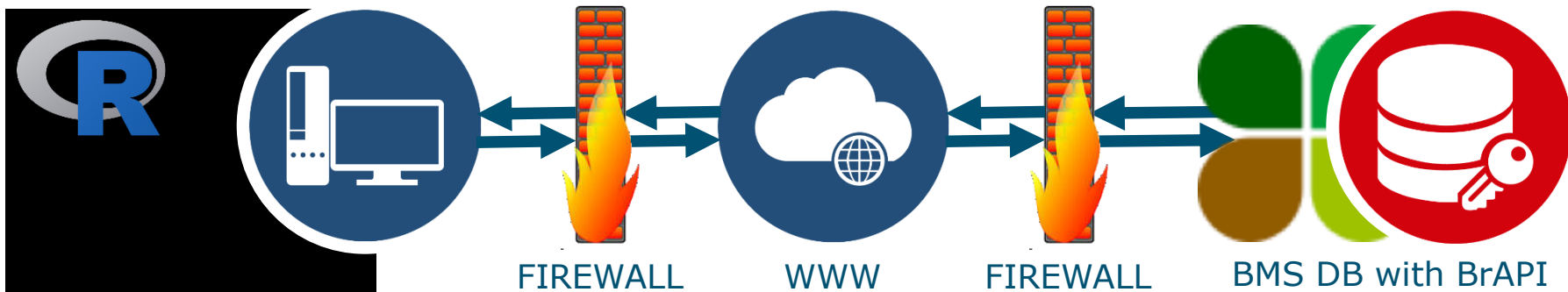
You are using version 1.2.1

Please register any issues at: <https://github.com/CIP-RIU/brapi/issues>

```
> # Create a connection object
```

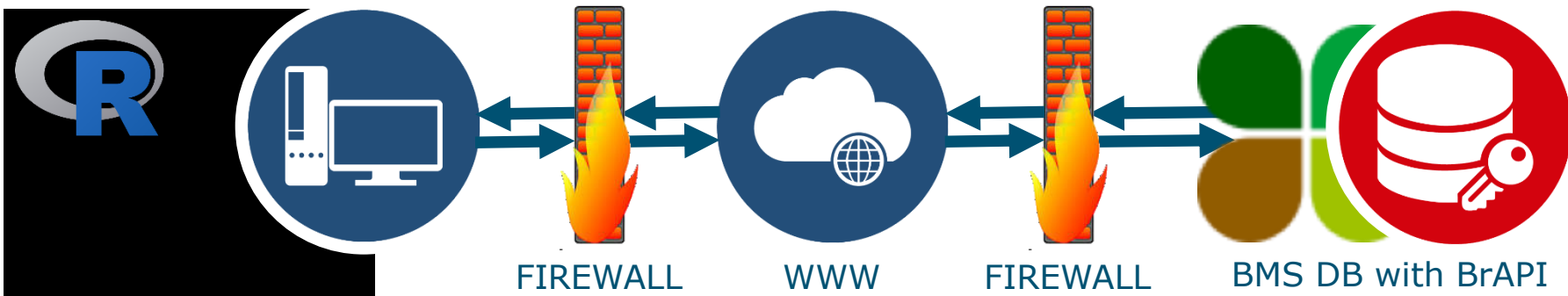
```
> bmscon <- ba_connect(brapiDb = null, secure = FALSE, protocol = "http://",  
+                      db = ..., port = ... # Hidden for security  
+                      apipath = "bmsapi", multicrop = TRUE, crop = "",  
+                      user = "...", password = "...", # Hidden for security  
+                      token = "", # Assigned at Authentication  
+                      bms = TRUE)
```

Example Software BrAPI R Package (II)



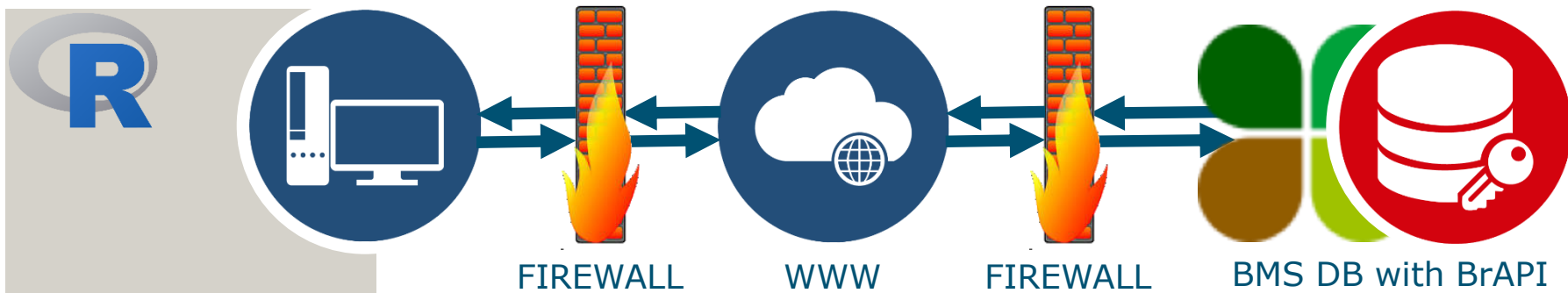
```
> bmscon <- ba_login(con = bmscon) # Returns the Authentication Token
Authenticated!
> (crops <- ba_crops(con = bmscon, rclass = "vector"))
[1] "rice" "wheat"
Attr(, "class")
[1] "character" "ba_crops"
> # Currently two crops on the BMS/BrAPI development server
> # BMS is a multi-crop DB
> # Assign "wheat" as the crop to be used
> bmscon$crop <- crops[2]
> # Extract Phenotypical data for two studies as table from BMS
> studyTable21 <- ba_studies_table(con = bmscon, studyDbId = "21")
> studyTable38 <- ba_studies_table(con = bmscon, studyDbId = "38")
```

Example Software BrAPI R Package (III)



```
> # Display the first three rows of the data extracted from BMS
> head(studyTable21, n = 3)[, c(1, 5, 7, 9, 10, 11)] # A few selected columns are shown
  year locationName germplasmName plotNumber replicate GY_Calc_tha|22661
1 2017      Site01      BELIKH02          1          1          3.957
2 2017      Site01      SAHEL77          2          1          4.222
3 2017      Site01      O.ZENADI          3          1          2.481
> head(studyTable38, n = 3)[, c(1, 5, 7, 9, 10, 11)] # A few selected columns are shown
  year locationName germplasmName plotNumber replicate GY_Calc_tha|22661
1 2017      Site18      BELIKH02          1          1          1.187
2 2017      Site18      HEBD/GDO          2          1          0.974
3 2017      Site18      O.ZENADI          3          1          1.887
```


Use Case BrAPI and gxe

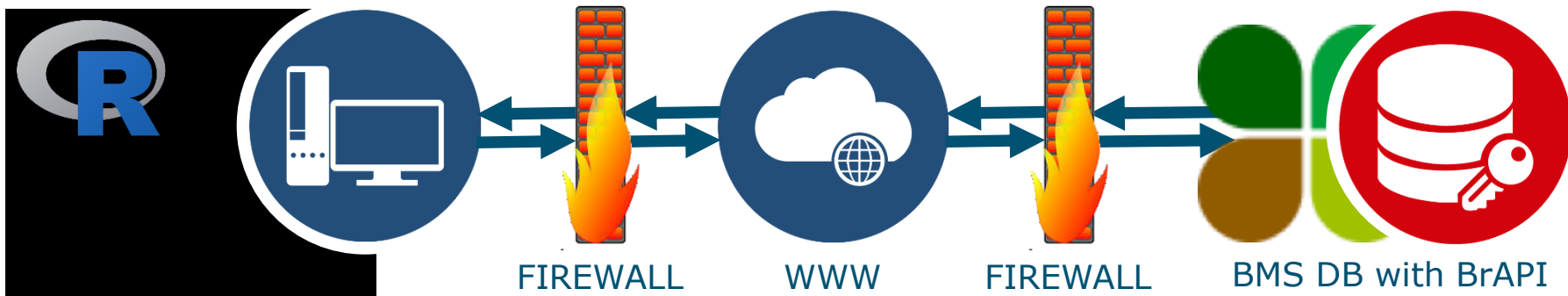


- Read data into R using BrAPI R package
- Analyse data using gxe R package

Advantages:

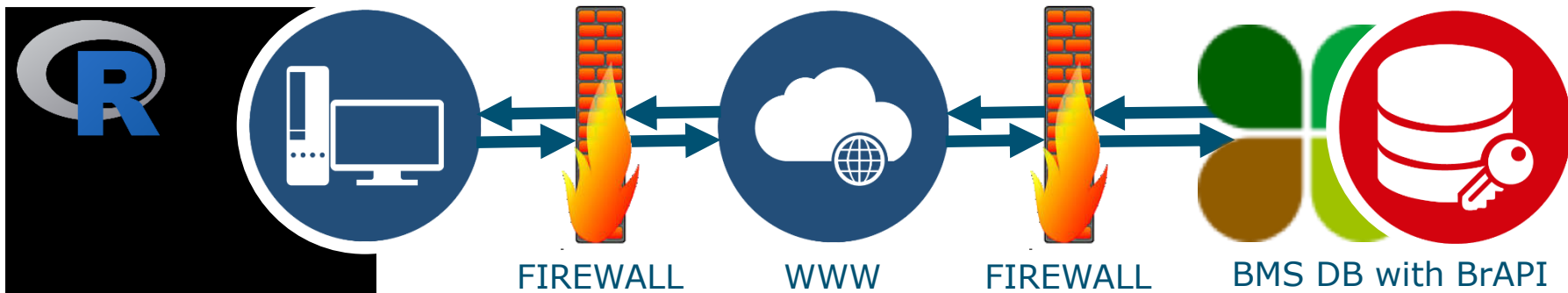
- Platform independent
- Open source
- Both packages are under development and will become available on CRAN

gxe Package (I)



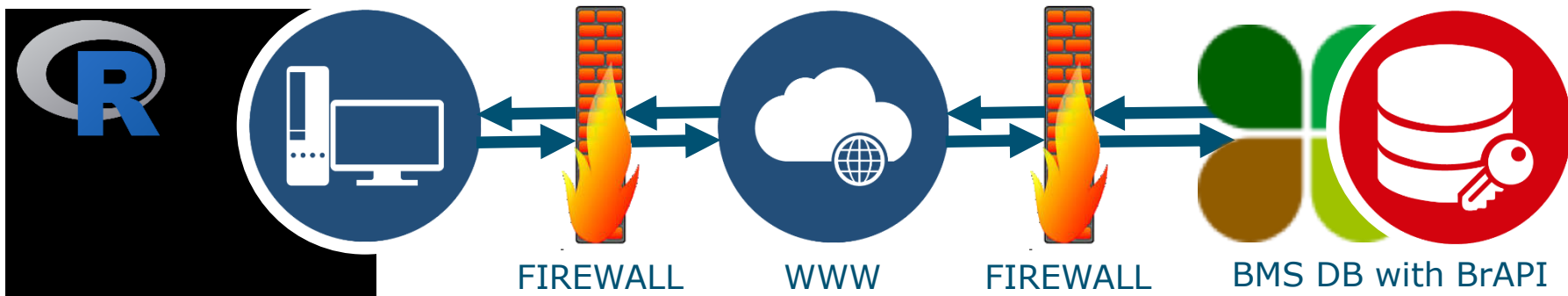
```
> # Load the gxe package into R
> library(gxe)
> # Create an object used for performing analysis
> wheatTD <- createTD(data = studyTableTot, genotype = "germplasmName",
                      env = "locationName", year = "year", repId = "replicate",
                      subBlock = "blockNumber", design = "rcbd")
> summary(wheatTD, traits = "GY_Calc_tha") # A summary of the data (partially displayed)
Summary statistics for GY_Calc_tha
  Number of observations 1344
  Number of missing values 0
           Mean 1.85
          Median 1.64
           Min 0.05
           Max 5.85
```

gxe Package (II)



```
> # Run Single Site Analysis using asreml
> SSA <- STRunModel(TD = wheatTD[wheatTD$env == "Site01", ], traits = GY_Calc_tha",
                    engine = "asreml")
> BLUPs <- STExtract(SSA, what = "BLUPs") # Extract BLUPs from output
> head(BLUPs, n = 3)
  genotype GY_Calc_tha
1  ARDENTE      3.641405
2  B.DUR194      3.625870
3  BELIKH02      3.597325
> (STExtract(SSA, what = "heritability")) # Extract heritability from output
  GY_Calc_tha
    0.7767329
```

gxe Package (III)



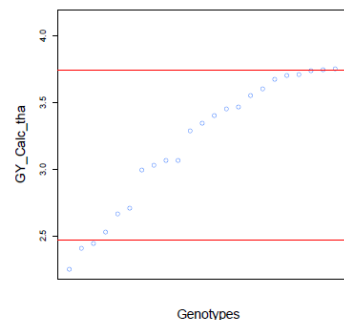
```
> report(SSA, outfile = "./reports/SSASite01.pdf") # Create a pdf and LaTeX report
```

List of best 10% genotypes with highest BLUES

genotype	GY_Calc_tha	SE
GTADUR	4.12	0.24
ARDENTE	3.76	0.24
CHENS	3.75	0.24

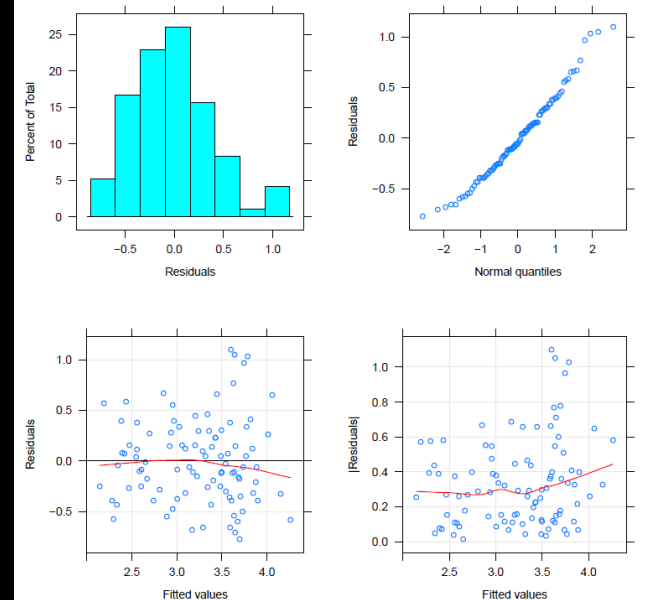
1 Background information

Modeling date 18-01-09
 Modeling time 11:35:28
 Modeling engine asreml
 Pipeline version 0.0.0.9000

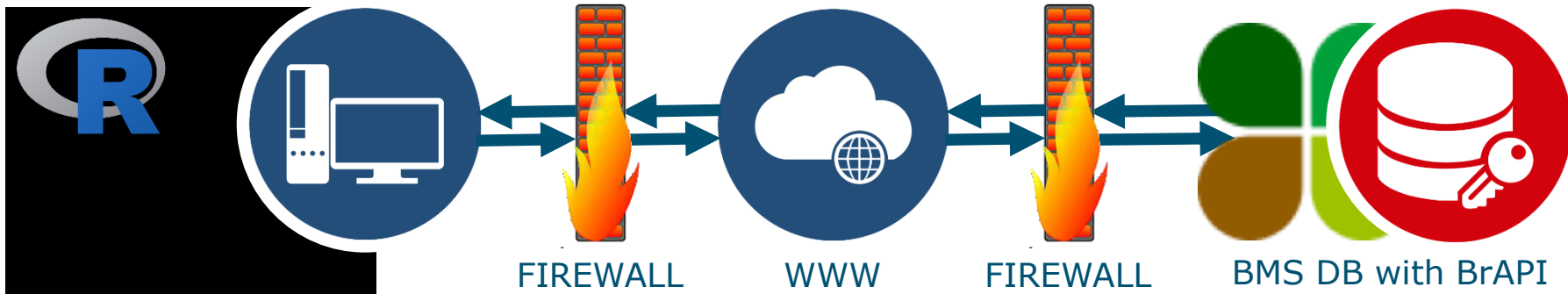


2 Description of the fitted model

Trait	GY_Calc_tha
Experimental design features	replicate (F)
Spatial model	id
Genotype	fixed



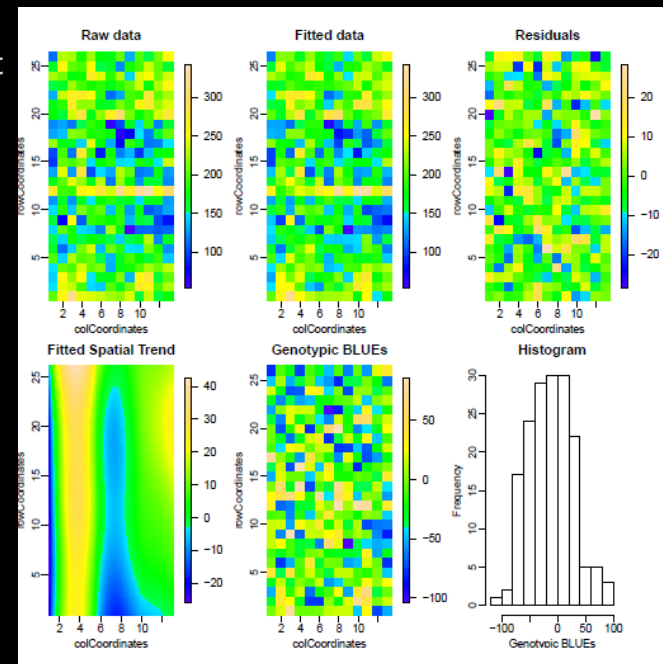
gxe Package (IV)



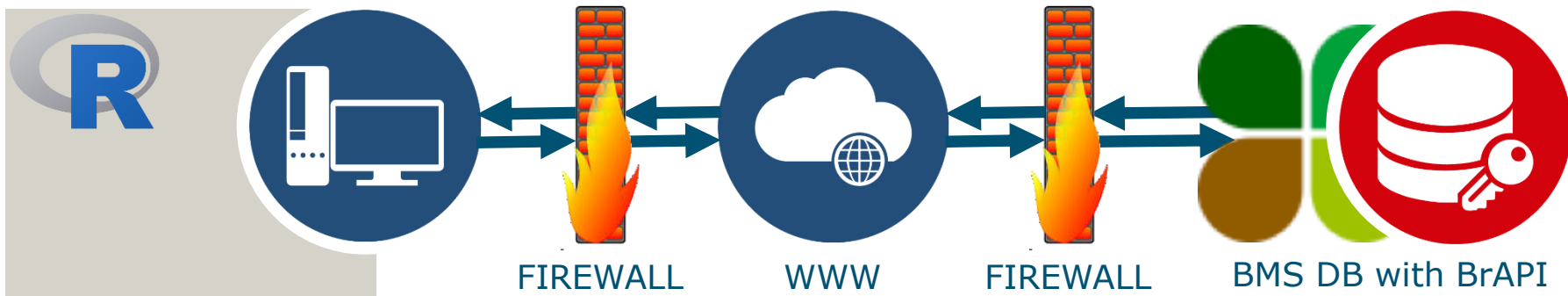
- > # For row column designs SpATS is used as default
- # to estimate the spatial trend.

Table 2: Effective dimensions

	Effective	Model	Nominal	Ratio	Type
Genotype	168.0	168	168	1.00	F
Intercept	1.0	1	1	1.00	F
Replicate	1.0	1	1	1.00	F
Row(replicate)	18.3	52	51	0.36	R
Col(replicate)	1.1	26	25	0.04	R
Linear trend along cols	1.0	1	1	1.00	S
Linear trend along rows	1.0	1	1	1.00	S
Linear trend along rows and cols	1.0	1	1	1.00	S
Smooth trend along cols	4.2	8	8	0.52	S
Smooth trend along rows	0.0	14	14	0.00	S
Linear trend in rows changing smoothly along cols	0.0	8	8	0.00	S
Linear trend in cols changing smoothly along rows	1.3	14	14	0.09	S
Smooth-by-smooth interaction trend over rows and cols	2.3	28	28	0.08	S
Total	200.2	323	321	0.62	
Residual	137.8				
Number of observations	338				



Use Case BrAPI and gxe R packages



Available functions gxe R package:

- Single Site Analysis: Using SpATS, asreml or lme4
- GxE analysis: Variance covariance modeling, AMMI, Finlay-Wilkinson, Stability
- QTL mapping: Quality checks, QTL Detection, Multi QTL modeling

Report generation available for all analyses:

- PDF and LaTeX
- Figures saved separately for reuse

Questions?

BrAPI:

- Unlocks data by making it FAIR

Internet:

- brapi.org links to BrAPI (API specs, events, etc.) and Software (a.o. BrAPI R package)

