

Implementation of an R Environments

Liam O'Suilleabhain

YHAT

February 04, 2019

Outline

Implementation
of an R
Environments

Liam
O'Suilleabhain

YHAT

Introduction

Environment

Application

Implementation

```
git clone https://github.com/losuilleabhain/AMORE.git
```

Introduction

Environment

Application

Implementation

Motivation

Implementation of an R Environments

YHAT

Introduction

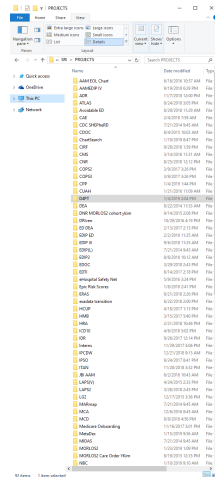
Reproducibility necessitates a **Standard Environment**

A Project has 3 sets of information

- 1 Resources
- 2 Operations

*A Functional Program
maps data to output*

- ### 3 Results



Many Shared Projects
Many Project Structures

?Startup

Implementation
of an R
Environments

Liam
O'Suilleabhain

YHAT

Introduction

Environment

Application

Implementation

Description:

In R, the startup mechanism is as follows.

Unless `'--no-envir'` was given on the command line, R searches for site and user files to process for setting environment variables. The name of the site file is the one pointed to by the environment variable `'R_ENVIRON'`; if this is unset, `'R_HOME/etc/Renviron.site'` is used (if it exists, which it does not in a 'factory-fresh' installation). The name of the user file can be specified by the `'R_ENVIRON_USER'` environment variable; if this is unset, the files searched for are `'.Renviron'` in the current or in the user's home directory (in that order). See 'Detailsr' for how the files are read.

Then R searches for the site-wide startup profile file of R code unless the command line option `'--no-site-file'` was given. The path of this file is taken from the value of the `'R_PROFILE'` environment variable (after tilde expansion). If this variable is unset, the default is `'R_HOME/etc/Rprofile.site'`, which is used if it exists (which it does not in a 'factory-fresh' installation). This code is sourced into the 'base' package. Users need to be careful not to unintentionally overwrite objects in 'base', and it is normally advisable to use 'local' if code needs to be executed: see the examples.

Then, unless `'--no-init-file'` was given, R searches for a user profile, a file of R code. The path of this file can be specified by the `'R_PROFILE_USER'` environment variable (and tilde expansion will be performed). If this is unset, a file called `'.Rprofile'` is searched for in the current directory or in the user's home directory (in that order). The user profile file is sourced into the workspace.

An R Environment

Implementation
of an R
Environments

Liam
O'Suilleabhain

YHAT

Introduction

Environment

Application

Implementation

1. R_ENVIRON

R_HOME/etc/Renviron.site

2. R_ENVIRON_USER

~/.Renviron

3. R_PROFILE

GROUP_HOME/Rprofile.site

4. R_PROFILE_USER

~/.Rprofile

Files in blue are available on github

One R Environment



Platform	Development	Notebooks	Datasets
Machine	DOR Desktop	KPIT Server	DOR Server
Processor	2.5GHz 4 Core	2.3GHz 8 Core	3.0GHz 32 Core
Memory	16GB	64GB	256GB
Storage	Unlimited	1TB	1TB
Support	DOR IT	KPIT	DOR IT

Example: Set Environment Variable

Implementation
of an R
Environments

Liam
O'Suilleabhain

YHAT

Introduction

Environment

Application

Implementation

Example `~/Renv` on Unix

```
R_LIBS="~/R/library"
```

Example `.Renv` on Windows

```
R_LIBS="C:/R/library"
```

R_ENVIRON - Global Environment (Optional)

Implementation
of an R
Environments

Liam
O'Suilleabhain

YHAT

Introduction

Environment

Application

Implementation

`$(R RHOME)/etc/RenvIRON.site`

Create a .RenvIRON file - `~/RenvIRON`

Set R_ENVIRON_USER

R_ENVIRON_USER - User Environment

Implementation
of an R
Environments

Liam
O'Suilleabhain

YHAT

Introduction

Environment

Application

Implementation

~/.Renviron

Create a group profile - GROUP_HOME/Rprofile.site

Set R_PROFILE

Create a shared library directory - GROUP_HOME/R_LIBS

Set R_LIBS_SITE

Create a personal library directory - ~/R_LIBS

Set R_LIBS_USER

R_PROFILE - Group Profile

Implementation
of an R
Environments

Liam
O'Suilleabhain

YHAT

Introduction

Environment

Application

Implementation

GROUP_HOME/Rprofile.site

Create .Rprofile - ~/ .Rprofile

Set R_PROFILE_USER

Create Project .Rprofiles

e.g. GROUP_HOME/RPROFILE/.RPROFILE

Create Function to source Project Environments

Define Renv

R_PROFILE_USER - User Profile (Freedom!)

Implementation
of an R
Environments

Liam
O'Suilleabhain

YHAT

Introduction

Environment

Application

Implementation

~/.Rprofile

Create Development Directory e.g.
~/Development/
Set Development Directory

Create Password Vault
~/pwv.txt
Load Passwords

A Project Profile

Implementation
of an R
Environments

Liam
O'Suilleabhain

YHAT

Introduction

Environment

Application

Implementation

GROUP_HOME/RPROFILE/Rprofile.RPROFILE

Create Resource Directory
Create Code Directory
Create Analysis Directory

Load Libraries
Set Project Options
Load Project Resources
Set Database Connections

Content Management

Implementation
of an R
Environments

Liam
O'Suilleabhain

YHAT

Introduction

Environment

Application

Implementation

The screenshot shows the D4PT web interface. At the top, there's a header with the D4PT logo and the text "Data Driven Determination of Dynamic Patient Trajectories (Project 01-01)". Below this, there's a navigation bar with links like "Star", "Fork", "HTTP", and a URL "http://kordy-gitlab.kaiser.org/151". There are also buttons for "Add README", "Add Changing", "Add License", "Add Contribution guide", "Add Kubernetes cluster", and "Set up CXC".

The main content area features a section titled "Auto Deploy" with a sub-header "It will automatically build, test, and deploy your application based on a predefined CXC configuration." Below this, there's a link to "Learn more in the Auto Deploy documentation" and a button to "Enable in settings".

Below the "Auto Deploy" section, there's a table with columns "Name", "Last commit", and "Last update". The table lists various files and their update status:

Name	Last commit	Last update
lymph_challenge	Updated	1 month ago
analysis_master_cohort1	Updated: rna	2 weeks ago
CODE	Updated: analysis data	1 week ago
NOTES	Notes accounting for discrepancy between number...	1 month ago
comments	Updated: code	1 month ago
AggrData-D4PT	Updated: network	2 weeks ago
gripense	Updated: gr	4 months ago
model_rna	Updated: modelling code	2 months ago
raw_phicapsseq_2018_1113raw	Updated: rna	2 weeks ago
reporter	Modified: RCC_rna	1 month ago

This PC > x225967 (l\pdongsub.kaiser.org) > rsch > srl > projects > D4PT > DATA			
Name	Date modified	Type	Size
ablapr_master_cohort1	1/16/2019 3:25 PM	File	5,198 KB
ablapr_master_cohort1	1/16/2019 3:44 PM	FF File	1 KB
advance_illness	1/17/2019 9:14 AM	FF File	1 KB
alibad_mnn	1/17/2019 9:13 AM	FF File	1 KB
analysis1_master_cohort1	1/16/2019 3:25 PM	File	51,950 KB
analysis1_master_cohort1	1/17/2019 11:33 AM	Microsoft Excel C...	2,955,727 KB
analysis3der_master_cohort1	1/17/2019 12:50 PM	Microsoft Excel C...	311,744 KB
analysis3win_master_cohort1	1/16/2019 4:41 PM	FF File	54,217 KB
bad_mnn	1/17/2019 9:13 AM	FF File	1 KB
bmi_master_cohort1	1/16/2019 3:25 PM	File	25,177 KB
bmi_master_cohort1	1/16/2019 3:45 PM	FF File	1 KB
copa2_master_cohort1	1/16/2019 3:38 PM	File	11,037 KB
copa2_master_cohort1	1/16/2019 4:07 PM	FF File	1 KB
CORR_MBN_PL3_NB	1/17/2019 9:13 AM	FF File	1 KB
cost_master_cohort1	1/4/2019 4:32 PM	File	9,953 KB
D4PT	1/17/2019 9:22 AM	FF File	65,061 KB
D4PTGP_COHORT1	1/16/2019 3:35 PM	File	28,241 KB
data_use	1/17/2019 9:13 AM	FF File	1 KB
death_post	1/17/2019 9:13 AM	FF File	1 KB
diccg_master_cohort1	1/16/2019 3:25 PM	File	46,631 KB
diccg_master_cohort1	1/16/2019 3:46 PM	FF File	1 KB
gp_post	1/17/2019 9:13 AM	FF File	1 KB
gp_pre	1/17/2019 9:13 AM	FF File	1 KB
hgatc_master_cohort1	1/16/2019 3:25 PM	File	6,553 KB
hgatc_master_cohort1	1/16/2019 3:47 PM	FF File	1 KB
hudi1_master_cohort1	1/16/2019 3:25 PM	File	18,865 KB
hudi1_master_cohort1	1/16/2019 3:50 PM	FF File	10,844 KB
master_mnn_cohort1	1/16/2019 3:31 PM	File	19,710 KB
medicare	1/17/2019 9:14 AM	FF File	1 KB
mortality_master_cohort1	1/16/2019 3:25 PM	File	12,294 KB
mortality_master_cohort1	1/16/2019 3:51 PM	FF File	10,844 KB
mnn_lookup_cohort1	1/16/2019 3:33 PM	File	20,847 KB
nam_gp	1/17/2019 9:14 AM	FF File	1 KB
nam_master_cohort1	1/16/2019 3:25 PM	File	34,431 KB
nam_master_cohort1	1/17/2019 9:19 AM	FF File	10,844 KB
phnrg_master_cohort1	1/16/2019 3:25 PM	File	912 KB
phnrg_master_cohort1	1/16/2019 3:51 PM	FF File	1 KB
rcc_master_cohort1	1/16/2019 3:25 PM	File	121,183 KB
rcc_master_cohort1	1/16/2019 4:06 PM	FF File	1 KB
nutl1_master_cohort1	1/16/2019 3:25 PM	File	35,863 KB
nutl1_master_cohort1	1/16/2019 4:09 PM	FF File	10,844 KB
ucda_master_cohort1	1/16/2019 3:25 PM	File	196,316 KB
ucda_master_cohort1	1/16/2019 4:17 PM	FF File	10,844 KB

Project Workflow

Implementation
of an R
Environments

Liam
O'Suilleabhain

YHAT

Introduction

Environment

Application

Implementation

Option 1:

Reference Files and Operations Relative to project path e.g.
`source("./CODE/analysis/models.r")`

Option 2:

Separate git controlled code from the project e.g.
`source(paste0(DATA,"analysis/models_output"))`

Suggestions

Implementation
of an R
Environments

Liam
O'Suilleabhain

YHAT

Introduction

Environment

Application

Implementation

Only manage code that creates data or analysis

Management

Content

Structure

Maintain same format for Code and Analysis directories

Maintenance

Operations - Code

Output - Analysis

Option 1 vs. Option 2

Option 1 is less verbose

Option 2 provides extra flexibility

Option 2 isolates code maintenance

Develop Packages for Healthcare Analytics

Maintain git repositories with open-source code

Acknowledgements

Implementation
of an R
Environments

Liam
O'Suilleabhain

YHAT

Introduction

Environment

Application

Implementation

Thanks:

SRI for supporting this framework.

Chris Paciorek for ideas and advice.

Raj, Alejandro, Gina and Brian for helpful conversations.