



# R and SAS working together

## DNR's experience moving to LSAF 5.3

Paul Vervuren, Danone Nutricia Research  
Jean-Marc Ferran, Qualiance



QUALIANCE

TT07 - PHUSE EU CONNECT 2021

# AGENDA

- The Business Case
- LSAF 5.3 Upgrade – A 3-step process
- R in LSAF – The DNR Approach
- Package Management
- User Experience Summary – R Studio vs. R in LSAF
- Conclusions

R and LSAF – Challenges and Solutions  
John Hendrickx  
PHUSE EMEA Autumn SDE 2021





QUALIANCE

*Thanks to **Jessie Zhao, Floor van Oudenhoven, John Hendrickx, Herman Ament** and **Daan van Nieuwenhuijzen** from the R subteam for the outstanding work in evaluating and implementing “R in LSAF” at Danone Nutricia Research.*

# R IN LSAF

## THE BUSINESS CASE

- 95% of Statisticians are proficient in R
- 30% of Statistical Programmers are trained in Statistics and use R
  - The remaining Statistical Programmers have recently been trained in R



### Internal LSAF survey

*Which area(s) would you still like to see improved or developed in future versions of LSAF? (Select at least 1, max 3)*

### Results for “R integration”:

|                         | 2020 (LSAF 4.7) | 2021 (LSAF 5.2) |
|-------------------------|-----------------|-----------------|
| Statistics              | 63%             | 80%             |
| Statistical Programming | 23%             | 12.5%           |
| Data Management         | 12%             | 0%              |

SAS® Life Science Analytics Framework - Repository

/SAS/LIPIDIET R TESTING/Files/60 Study Report/50 Output Programs

Repository

- SAS
  - 000 Sample Size...
  - 001 Sample Pro...
  - 999 E2E Tenuto ...
  - 999 Test server t...
  - 999 UAT LSAF 5...
  - LIPIDIET R TES...
    - Files
      - 00 Docum...
      - 01 Study ...
      - 09 STAT S...
      - 20 Data R...
      - 60 Study ...
        - 00 Doc...
        - 01 Mac...
        - 02 Met...
        - 03 For...
        - 04 Imp...

Text to filter 35 items

|                          | ↑ | Name                | ↑ | Size     | Modified By    | Date Modified       | S... | Reposit... | Worksp... |
|--------------------------|---|---------------------|---|----------|----------------|---------------------|------|------------|-----------|
| <input type="checkbox"/> |   | Job                 |   |          | John Hendrickx | Sep 27, 2021, 1:... |      |            |           |
| <input type="checkbox"/> |   | Listing             |   |          | John Hendrickx | May 26, 2021, 1...  |      |            |           |
| <input type="checkbox"/> |   | Log                 |   |          | John Hendrickx | Sep 27, 2021, 1:... |      |            |           |
| <input type="checkbox"/> |   | Manifest            |   |          | John Hendrickx | Sep 27, 2021, 1:... |      |            |           |
| <input type="checkbox"/> |   | A01_Joint_model_... |   | 13.18... | John Hendrickx | Sep 17, 2021, 4:... |      | -          |           |
| <input type="checkbox"/> |   | A01_Joint_model_... |   | 12.94... | John Hendrickx | Sep 27, 2021, 1:... |      | -          | ⚠ -       |
| <input type="checkbox"/> |   | A01_Joint_model_... |   | 13.06... | John Hendrickx | May 26, 2021, 1...  |      |            |           |
| <input type="checkbox"/> |   | A01_Joint_model_... |   | 53.64... | John Hendrickx | Jun 28, 2021, 4:... |      |            |           |
| <input type="checkbox"/> |   | A01_Joint_model_... |   | 13.09... | John Hendrickx | May 26, 2021, 8:... |      |            |           |
| <input type="checkbox"/> |   | A01_Joint_model_... |   | 13.06... | John Hendrickx | May 26, 2021, 8:... |      |            |           |

SAS® Life Science Analytics Framework - Sessions

R Session: 1

Workspace

- SAS
  - 00 Rlibrary
  - Files
  - LIPIDIET R TEST...
    - Files
      - 20 Data Re...
      - 60 Study R...
        - 00 Doc...
        - 01 Macros
        - 02 Meta...
        - 03 For...
        - 04 Impo...
        - 40 Ada...
        - 41 Ada...
        - 50 Outp...
        - 55 Outp...
        - 60 Publi...

A01\_Joint model\_3years\_HIPPVU\_UATR.R

Code Log

```

46 #####
47 ### HIPPVU ###
48 #####
49
50 nutricia_config()
51
52 # Check variables in _overhead.R are sourced
53 print(study_name)
54 print(study_id)
55 print(Viedoc_dt)
56
57 # load and prepare data
58 adqs_nutricia <- read_sas("R:/SAS/LIPIDIET/Files/70 Post-hoc Analysis/05_LDD EXT1 Study Report Analysis/41 ADaM Data/adqs.
59 adqs_nutricia <- read_sas(getAdam("adqs"))
60 head(adqs_nutricia)
61
62 # read CSV and XLSX files
63 biostatsCSV <- read.csv(getAdam("biostats",ext=".csv"))
64 head(biostatsCSV)
65
66 #biostatsXLSX <- read.xlsx(getAdam("biostats",ext=".xlsx"))
67 #head(biostatsXLSX)
68
69
70

```

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Project: (None)

Environment History Connecto

Files Plots Packages

Install Update

System Library

Files

```

46 #####
47 ### HIPVL ###
48 #####
49
50 nutricia_config()
51
52 # Check variables in _overhead.R are sourced
53 print(study_name)
54 print(study_id)
55 print(Viedoc_dt)
56
57 # load and prepare data
58 # adqs_nutricia <- read_sas("R:/SAS/LIPIDIET/Files/70 Post-hoc Analysis/05_LDD
59 adqs_nutricia <- read_sas(getAdam("adqs"))
60 head(adqs_nutricia)
61
62 # read CSV and XLSX files
63 biostatsCSV <- read.csv(getAdam("biostats",ext=".csv"))
64 head(biostatsCSV)
65
66

```

Console

SAS® Drug Development Desktop Connection

Volumes Preferences

| Volume name                | Status           | Server URL                        |
|----------------------------|------------------|-----------------------------------|
| LSAF DNR Repository        | Disconnected     | https://xadprodtemp.ondemand....  |
| LSAF DNR Repository (TEST) | Disconnected     | https://xadtest.ondemand.sas.c... |
| LSAF DNR Workspace         | Disconnected     | https://xadprodtemp.ondemand....  |
| LSAF DNR Workspace (TEST)  | Connected as D:\ | https://xadtest.ondemand.sas.c... |

Configure... Disconnect

60 Study Report

File Home Share View

Clipboard Organize New Open Select

Pin to Quick access Copy Paste Cut Copy path Paste shortcut Move to Copy to Delete Rename New folder Easy access Properties History Edit Select all Select none Invert selection

This PC > LSAF DNR Workspace (TEST) (\SDDDC) (D:) > SAS > LIPIDIET R TESTING > Files > 60 Study Report

| Name                | Date modified      | Type        |
|---------------------|--------------------|-------------|
| 00 Documentation    | 8/24/2021 5:10 PM  | File folder |
| 01 Macros           | 9/21/2021 1:37 PM  | File folder |
| 02 Metadata         | 10/4/2021 9:08 AM  | File folder |
| 03 Formats          | 8/24/2021 5:10 PM  | File folder |
| 04 Import Data      | 8/24/2021 5:10 PM  | File folder |
| 40 ADaM Programs    | 8/24/2021 5:10 PM  | File folder |
| 41 ADaM Data        | 9/17/2021 2:22 PM  | File folder |
| 50 Output Programs  | 10/5/2021 10:02 AM | File folder |
| 55 Output Results   | 8/24/2021 5:10 PM  | File folder |
| 60 Publish Programs | 8/24/2021 5:10 PM  | File folder |
| 65 Publish Results  | 8/24/2021 5:10 PM  | File folder |
| 70 QC               | 8/24/2021 5:10 PM  | File folder |
| _overhead.R         | 9/17/2021 4:58 PM  | R File      |

# R IN LSAF

## THE BUSINESS CASE

### AS-IS

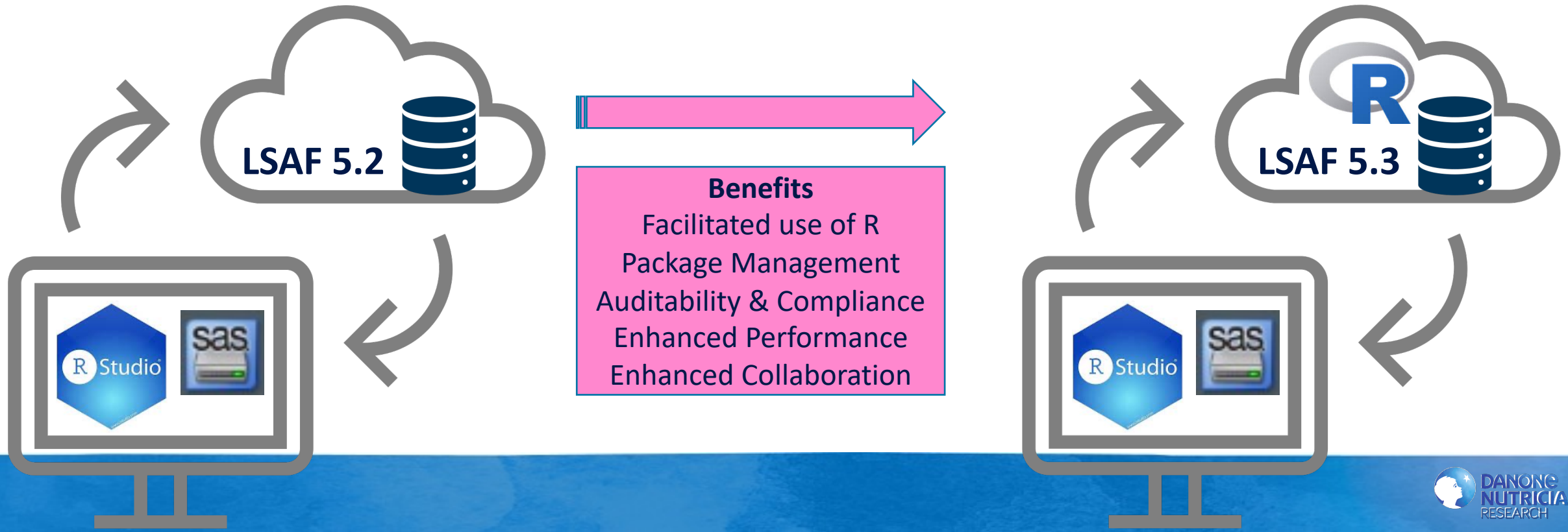
User-friendly interface with R Studio, but with risks & issues:

- Package use and auditability
- Performance issues through usage of SDD-DC
- Data fragmentation

### TO-BE

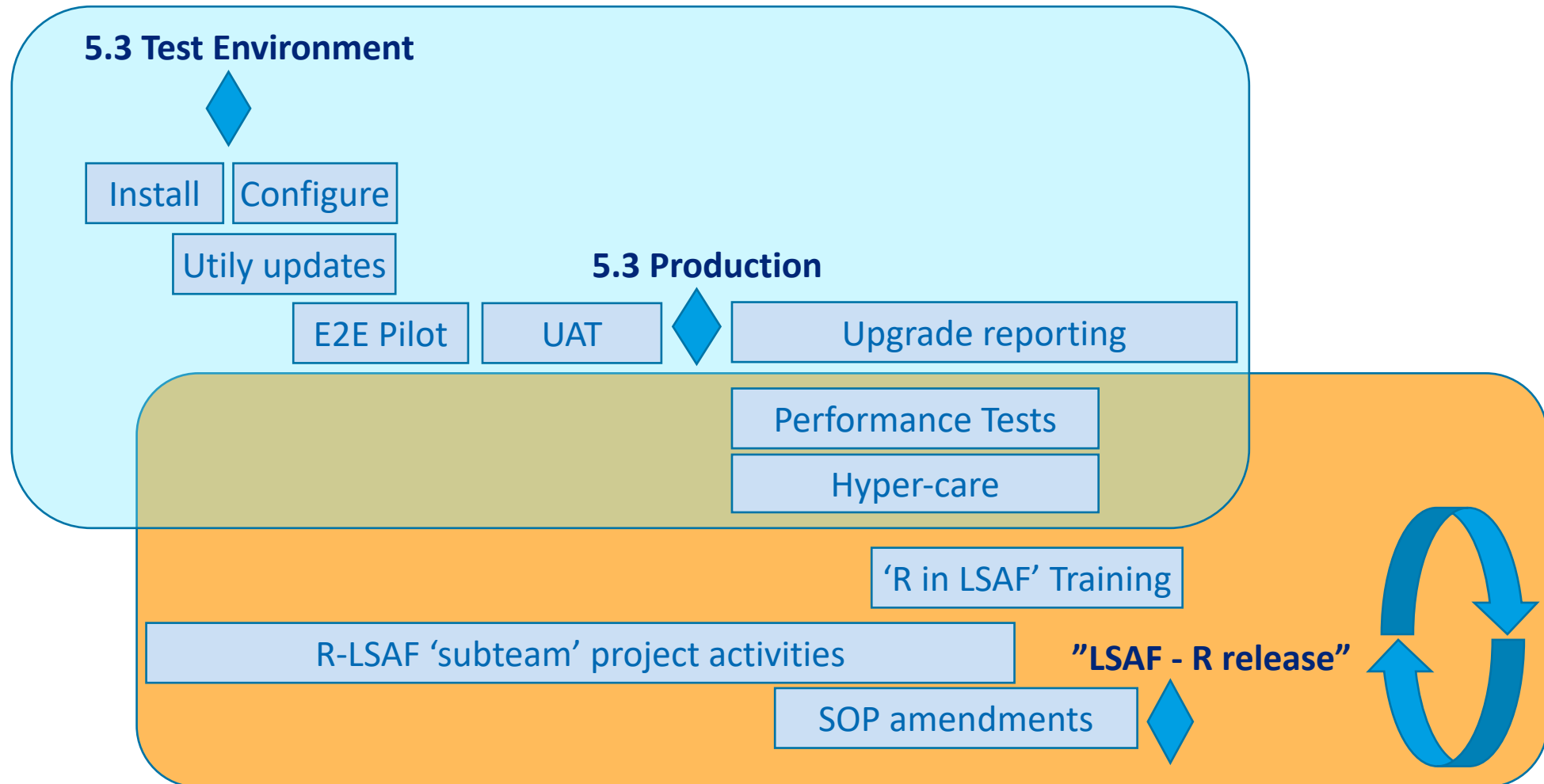
Hybrid solution

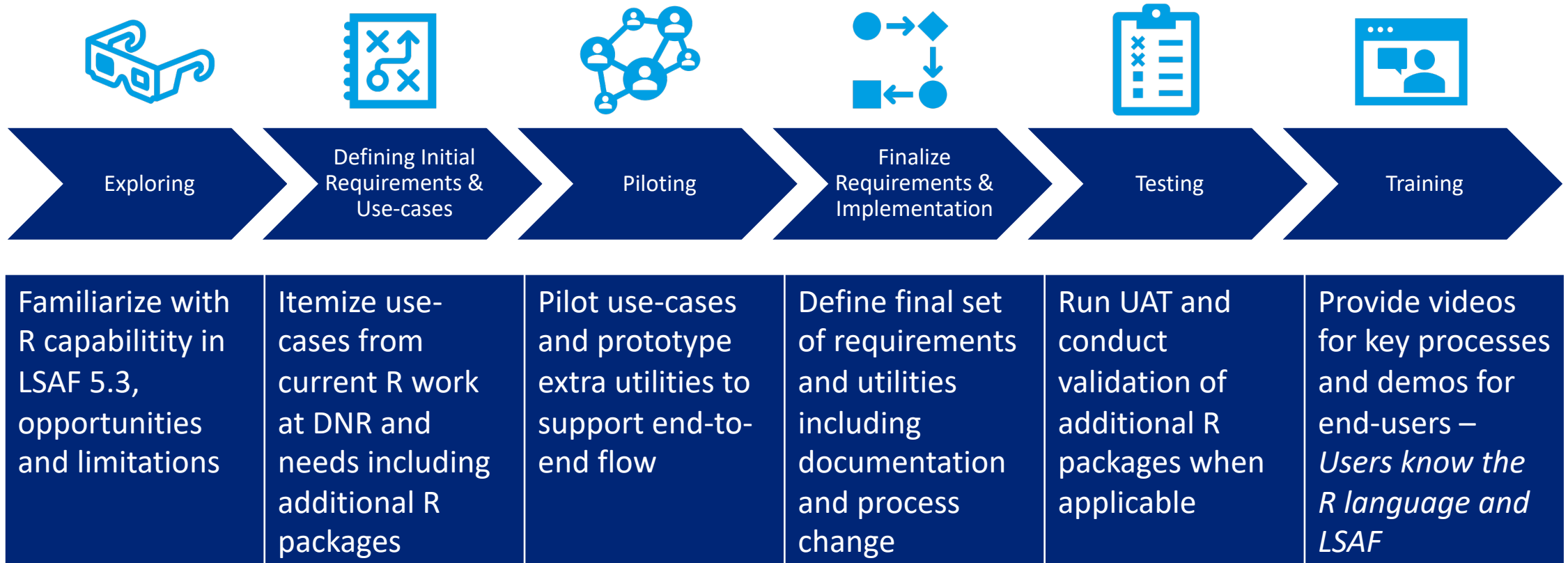
- R Studio for Development
- Controls in LSAF: packages, execution, audit trail
- Enhanced Performance



# APPROACH & PHASING

Q1 2021  Q3 2021



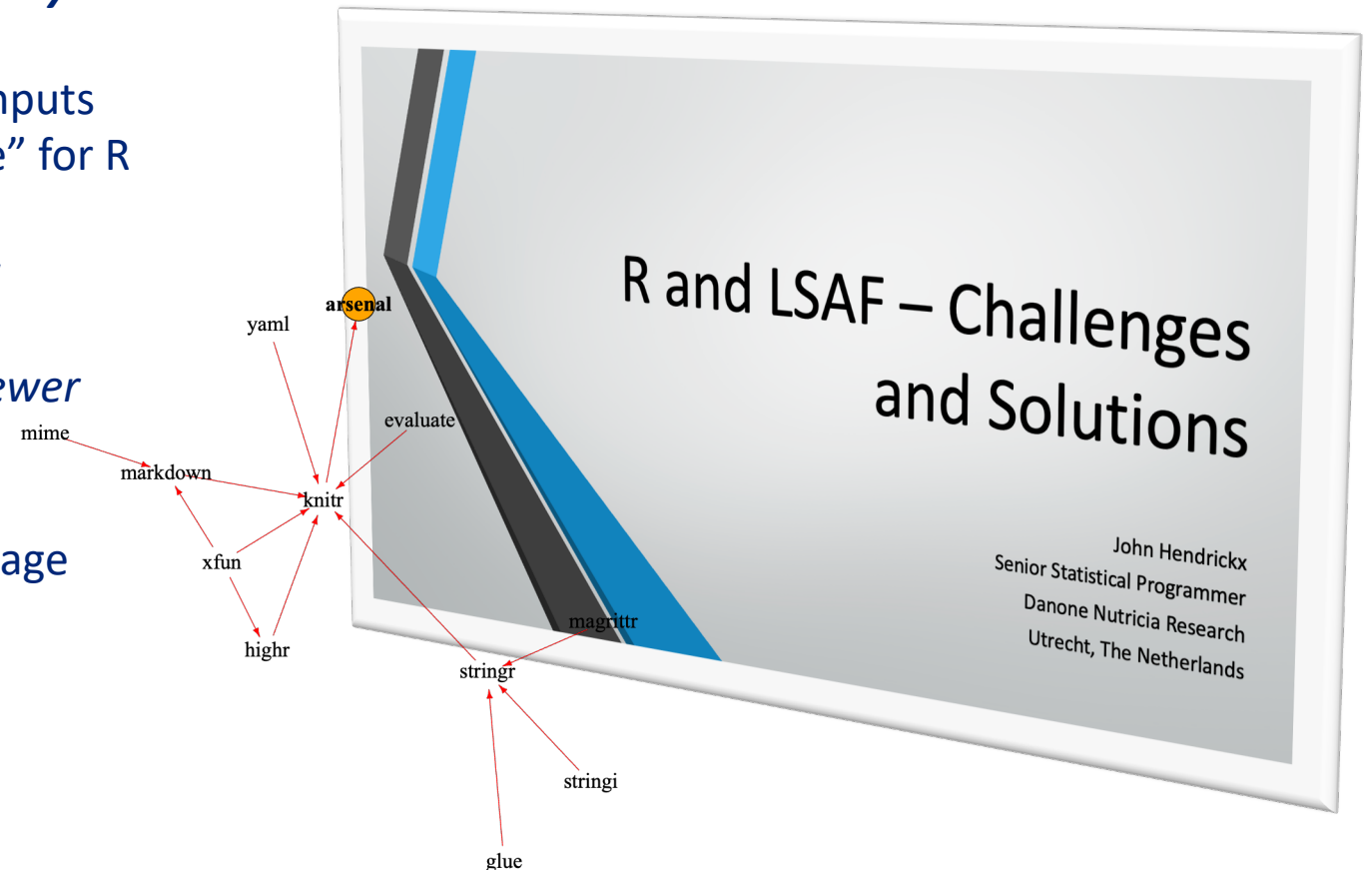


- **R Access function** – aka “nutricia\_config” with standard functions for reading / writing in standard locations and setting up study-specific variables *in LSAF and from R Studio seamlessly*
- **Job creation tool** for R to handle standard inputs and outputs in the lack of a “run & populate” for R
- **Install of “Pandoc”/“DT::datatable”** to view .Rdata datasets as HTML files
  - *Not needed in LSAF 5.4...DataFrame viewer*
- **Usage of MiniCRAN** to establish process to automate handling of dependencies in package installs within LSAF
- Etc.

## R and LSAF – Challenges and Solutions

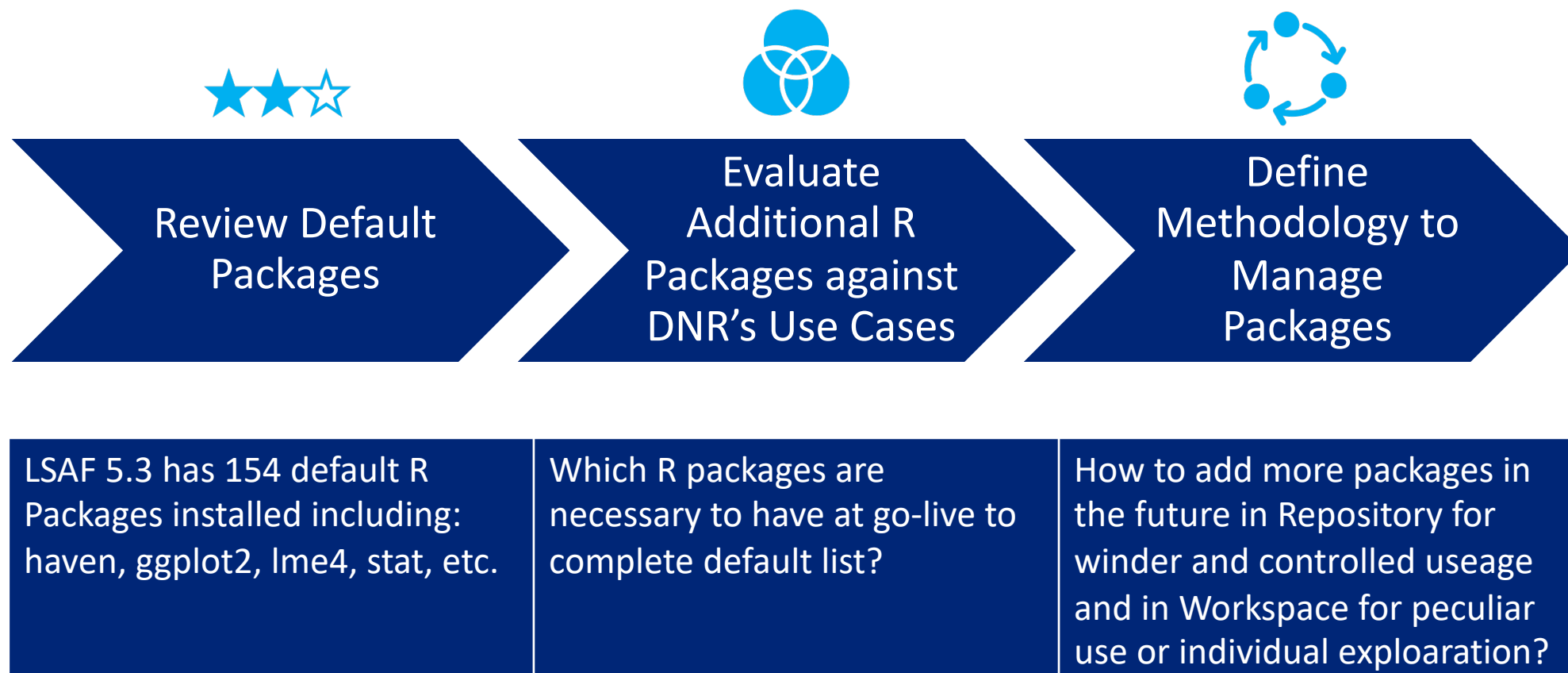
John Hendrickx

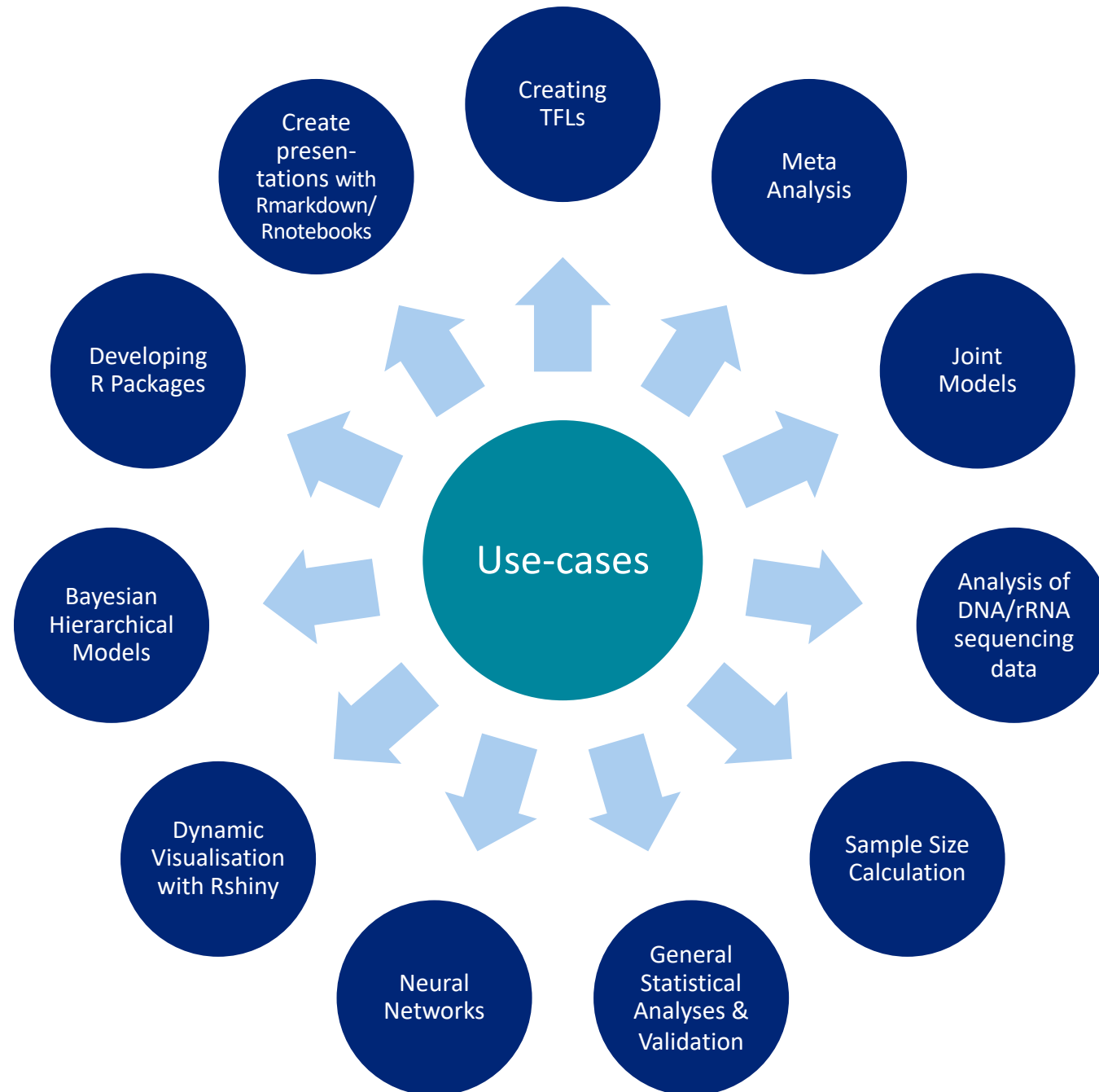
PHUSE EMEA Autumn SDE 2021



# PACKAGE MANAGEMENT

FINDING THE BALANCE – TRADE-OFF BETWEEN COMPLETENESS AND CONTROL





# PACKAGE MANAGEMENT

A RISK-BASED APPROACH FOR ASSESSING R PACKAGE ACCURACY WITHIN A VALIDATED INFRASTRUCTURE (PHARMAR.ORG)

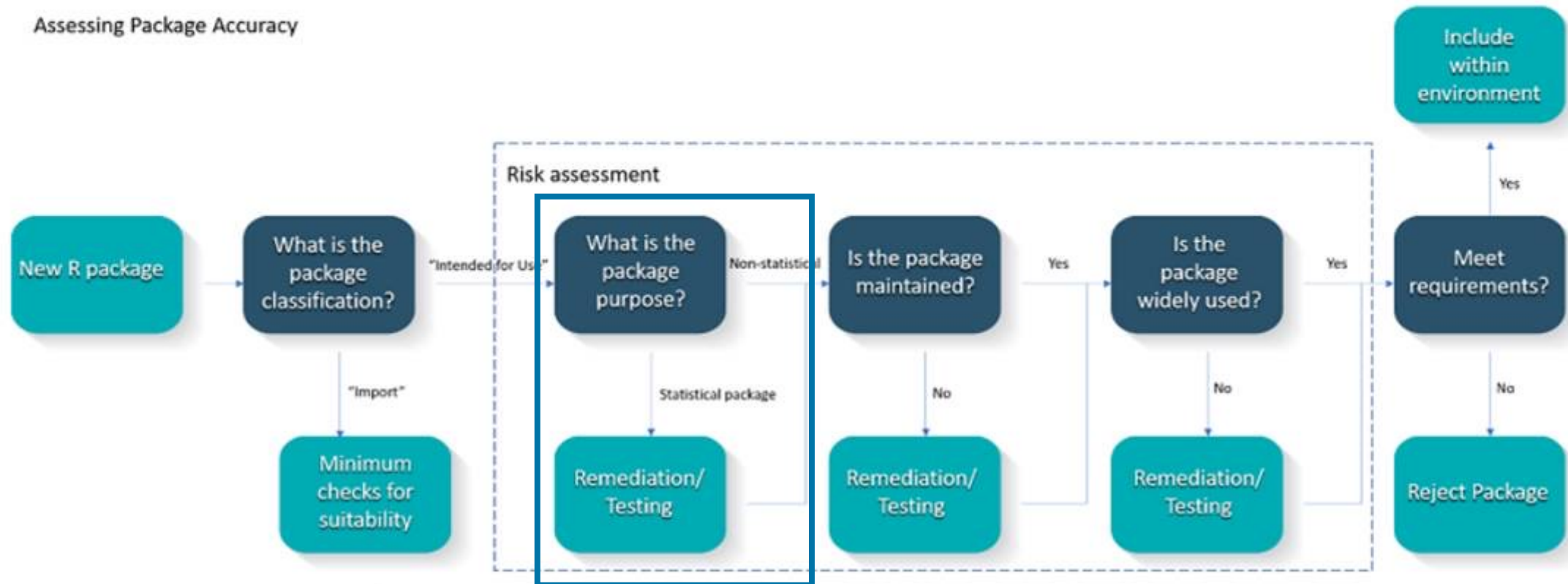
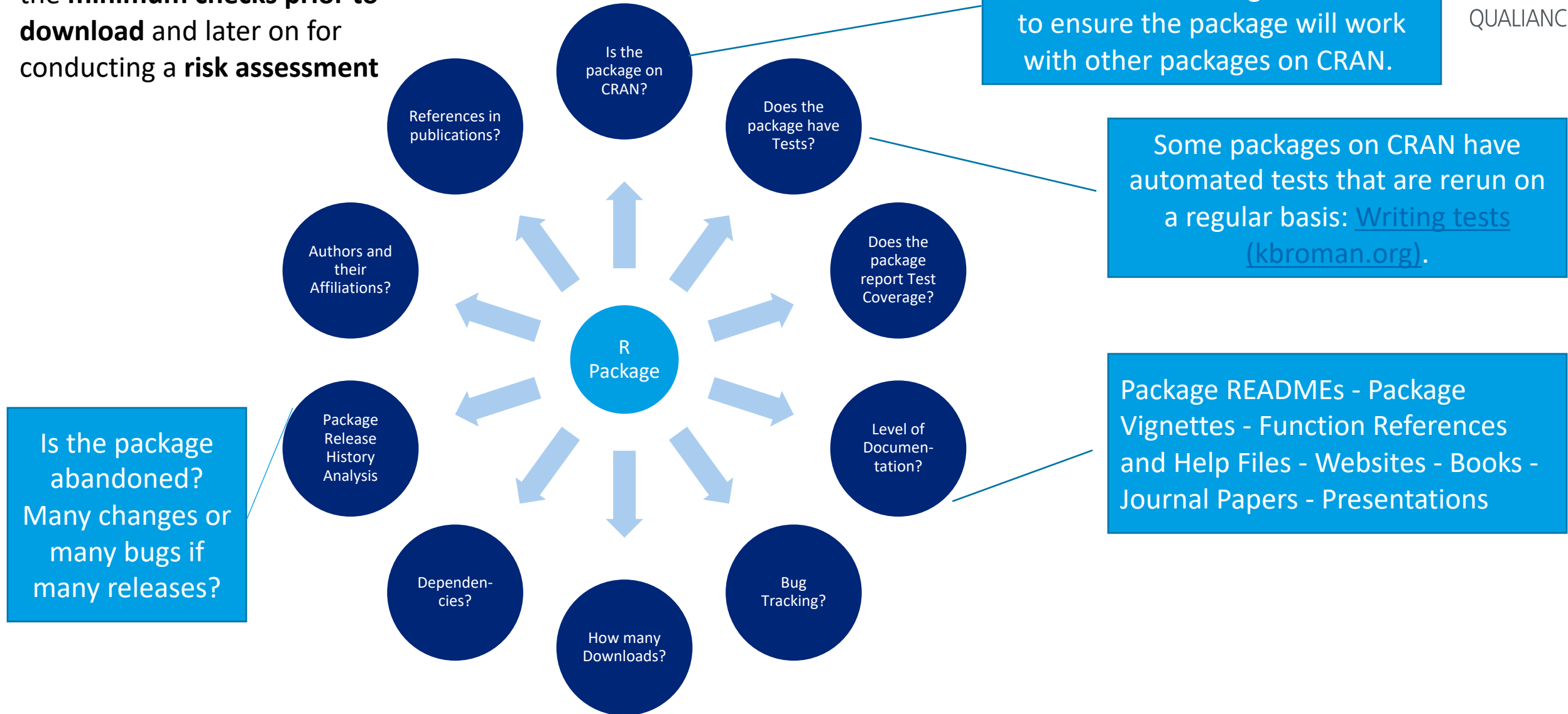


Figure 2 - Proposed Validation Workflow

More criteria to consider for the **minimum checks prior to download** and later on for conducting a **risk assessment**

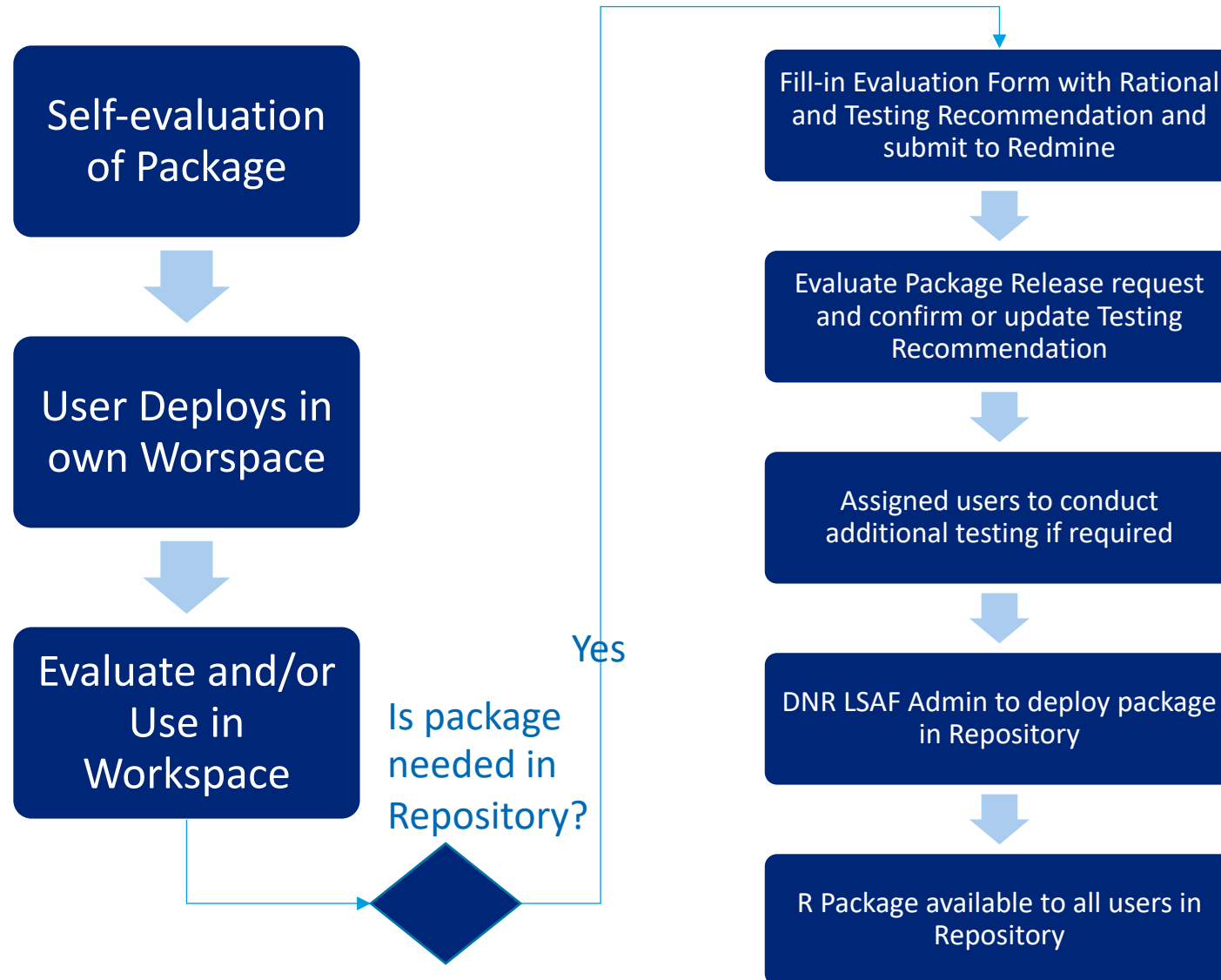


Source: [Environments: Validation \(rstudio.com\)](https://rstudio.com)

Other: Are there License Restrictions?

# PROCESS TO DEPLOY R PACKAGES IN WORKSPACE AND REPOSITORY

## HIGH-LEVEL PROCESS



# USER EXPERIENCE

## R IN LSAF VS. R STUDIO – A HYBRID APPROACH

| Feature              | R Studio                                                               | R in LSAF 5.3                                                                                                        |
|----------------------|------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------|
| Package Management   | No Control                                                             | Controlled Process in Repository                                                                                     |
| Package Dependencies | Fully Automated                                                        | 3-step Iterative Process                                                                                             |
| Use of .Rprofile     | Embedded                                                               | Alternative via a central R file in each study milestone                                                             |
| Program Editor       | Advanced                                                               | Decent                                                                                                               |
| Data Explorer        | Available for .Rdata files and other                                   | Alternative via use of “Pandoc”/“DT::datatable”<br>Possible to view csv files<br><b>DataFrame viewer in LSAF 5.4</b> |
| Log                  | Manual via command line                                                | Embedded                                                                                                             |
| Job                  | Supported                                                              | R Job supported                                                                                                      |
| Traceability         | Limited                                                                | Enabled through job manifests                                                                                        |
| Performance          | Limited by users’ laptops and use of Desktop Connection to access data | Enhanced through usage of LSAF cores and I/O directly in LSAF                                                        |

- Possibility to use R Studio with Desktop Connection for Development while final tuning can be done in LSAF

- R in LSAF is expected to bring significant benefits and will enhance the usage of R across Statistical Programming and Statistics, aside from and complementary to SAS.
- Compliance, Performance and Traceability are the main benefits of using R in LSAF.
- Possibility to work seamlessly on code from R Studio with Desktop Connection to read/write to LSAF Repository and Workspace and within LSAF only ensures we get the “best of both world”.
- LSAF 5.4 has a .Rdata Data Explorer planned for, which will provide a more complete Programming Environment.