

Single Day Event

Taku Sakaue, Tomoyuki Namai
Chugai Pharmaceutical Co., LTD.

Introduction to an example of R usage



Contents

- What's OSS
- OSS Pros and Cons
- Issues using OSS
- Introduce analysis environment
- Our activities for OSS



What's OSS

- What's OSS
 - OSS (Open Source Software) is computer software that is released under a license.

Grants user to **use**, **change**, and **distribute** the software and its source code to **anyone** and for **any purpose**.





OSS Pros and Cons

- **Pros** from viewpoint of pharmaceutical industry
 - Environment
 - Easy to get knowledge (sample program, TIPs, and troubleshooting)
 - Talent
 - Recruitment
 - Lot of people who learned R / Python in university. Less investment for education of around programming skill.
 - Easy to find effective player around data science



OSS Pros and Cons

- **Pros** from viewpoint of pharmaceutical industry
 - Cost
 - Suppress cost (free license) for implementation
 - Well compatibility of cloud environments
 - Independent of machine spec (scale up/down)
 - Usability
 - Variety module/package for graphic
 - Well-supported by ISE (Rstudio, Jupyter)



OSS Pros and Cons

- **Cons** from viewpoint of pharmaceutical industry
 - Environment
 - Quick lifecycle (need to keep multi version)
 - High frequency of software update
 - Validation issue
 - Implement/Maintain OSS derived software, tools, and analysis environment to observe GxP
 - » Required lot of human resources and expert of OSS and validation



OSS Pros and Cons

- **Cons** from viewpoint of pharmaceutical industry
 - Talent
 - Required support to existed SAS user (R/Python training)
 - Hard to shift R/Python from SAS (especially pure SAS programmer) due to uniqueness of SAS language
 - Cost
 - Increase maintenance cost
 - Maintain software version, tools, and environment by in-house expert

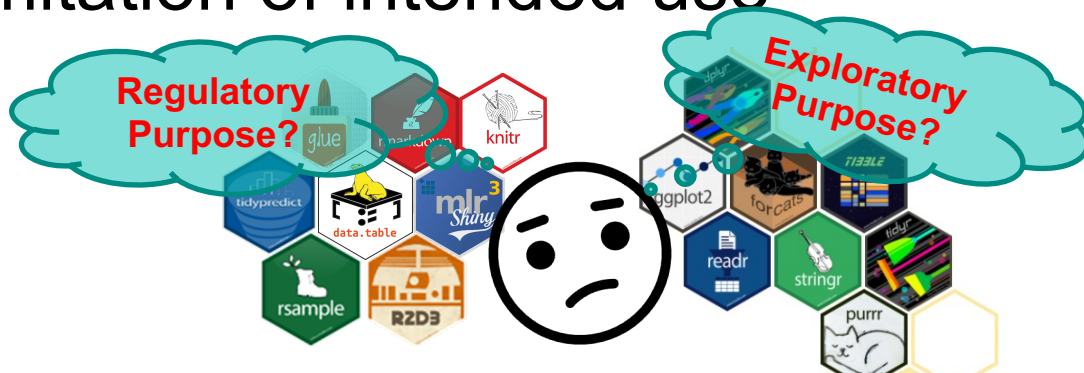


OSS Pros and Cons

- **Cons** from viewpoint of pharmaceutical industry
 - Usability
 - Not good at big data handling
 - Need to combine some technology to improve processing speed for big data handling

Issues using OSS

- Limitation of intended use



The credibility of the numerical results of the analysis depends on the quality and validity of the methods and software (both internally and externally written) used both for data management (data entry, storage, verification, correction and retrieval) and also for processing the data statistically. Data management activities should therefore be based on thorough and effective standard operating procedures. **The computer software used for data management and statistical analysis should be reliable, and documentation of appropriate software testing procedures should be available.**

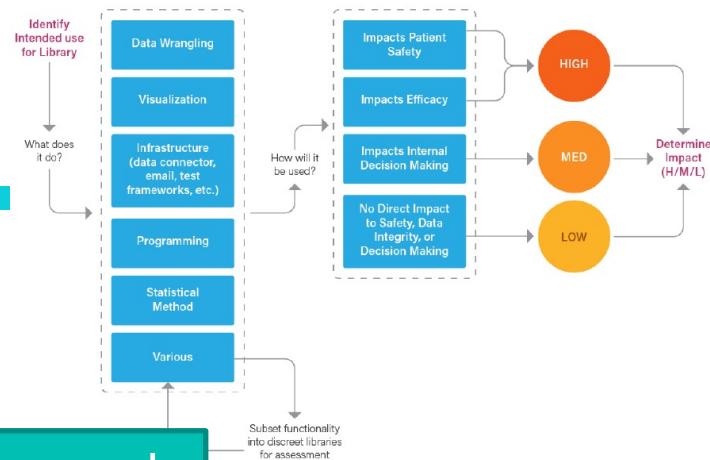


Issues using OSS

- How maintain reliability?



Modernization of statistical analytics (MSA) frame work



Confidence in Accuracy of Software Library

Low Impact High Confidence	Medium Impact High Confidence	High Impact High Confidence
Low Impact Medium Confidence	Medium Impact Medium Confidence	High Impact Medium Confidence
Low Impact Low Confidence	Medium Impact Low Confidence	High Impact Low Confidence

Impact of Software Library's Use for Regulatory Submissions

■ No/minimal additional testing ■ Limited Testing for Intended Use ■ Broad Testing



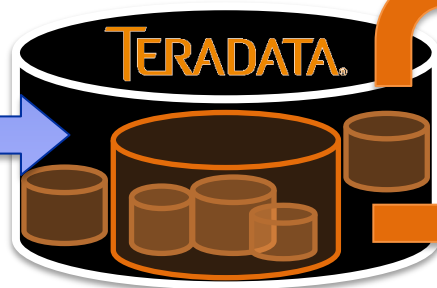
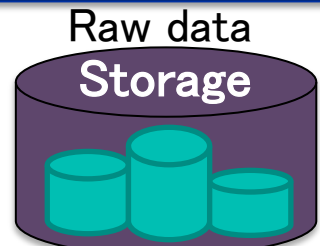


Introduce Analysis Environment

data management

data handling

data analysis



Confidentiality data
Analysis environment



R/Python
Analysis environment

SAS studio



SAS
Analysis environment



Introduce Analysis Environment

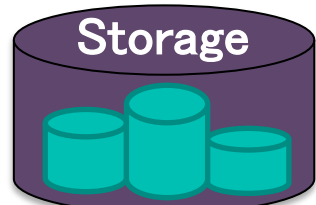
data management

data handling

data analysis

Raw data

Storage



Confidentiality data
Analysis environment



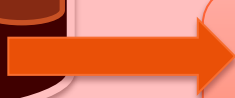
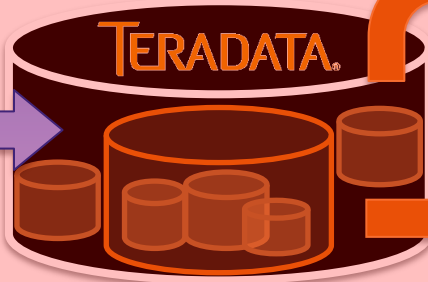
R/Python
Analysis environment

SAS studio



SAS
Analysis environment

TERADATA



Validated



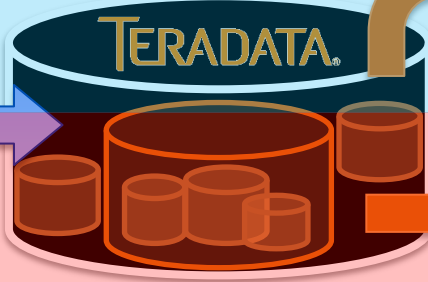
Introduce Analysis Environment

data management

data handling

data analysis

Raw data
Storage



Validated



Confidentiality data
Analysis environment

For exploratory analysis



R Studio
jupyter

Python
Analysis environment

SAS



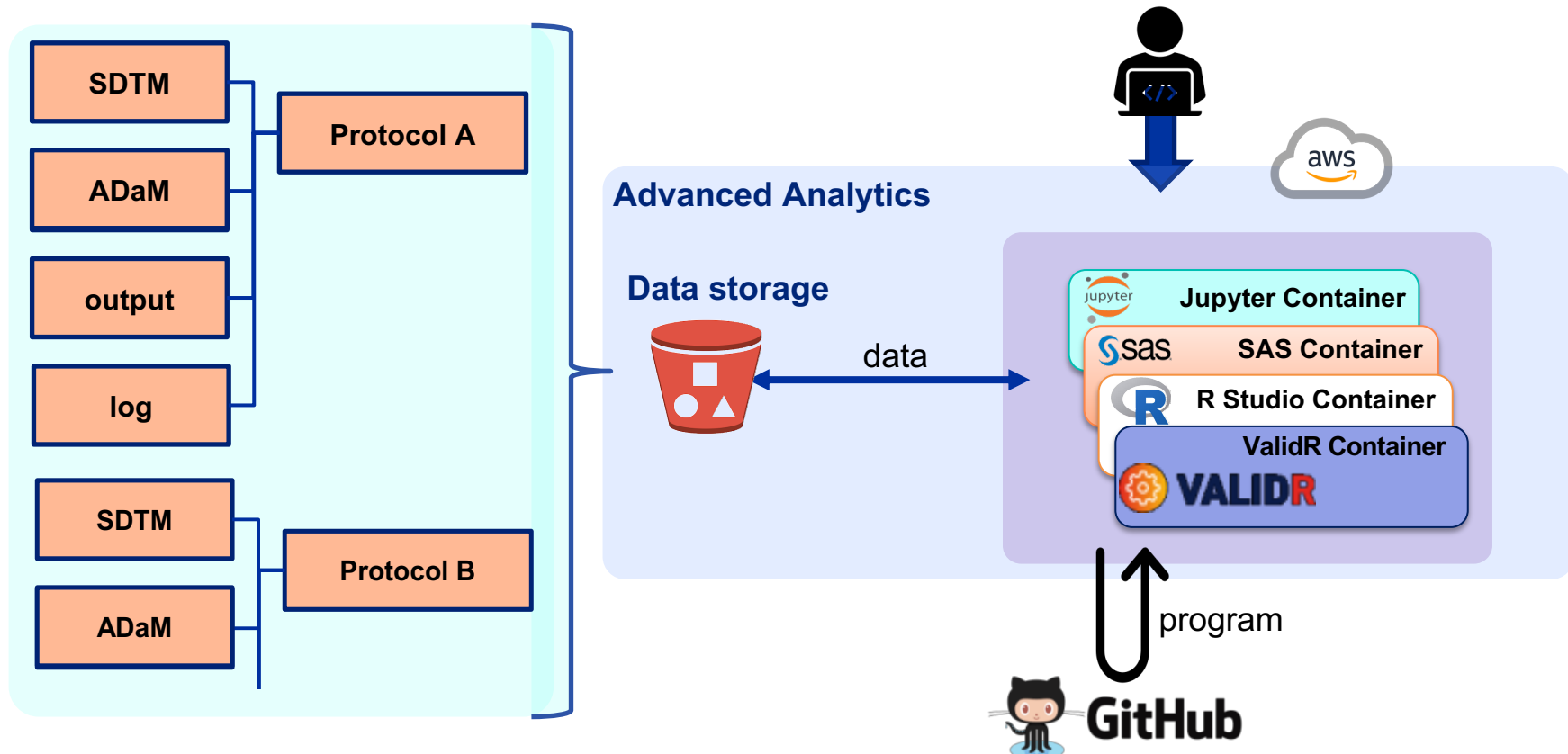
For regulatory analysis

SAS

SAS Analysis environment



Introduce Analysis Environment





Our Activities for OSS

- R training for SAS programmer
 - Prepared 4 months training course (Accel2R)
 - Side by side SAS to R learning
 - Using clinical data (SDTM) for exercise
 - Kindly learning method (Video > exercise (hands on) > test)
 - Group learning

Conclusion

- For reduction risk of accuracy of software library
 - Share use case of software library
 - Intended use
 - Validation status
 - Test case
 - Evidence of successful use

Collect, maintain, and share these information



Thank you



**The Global Healthcare
Data Science Community**

Contact Channels

📞 UK +44 1843 609600
📞 USA +1 609 514 5105
✉️ office@phuse.global
🌐 phuse.global

Social Media

🐦 [@phusetwitta](https://twitter.com/phusetwitta)
📘 [/phusebook](https://facebook.com/phusebook)
▶️ [/phusetube](https://youtube.com/phusetube)
🌐 [/company/phuse](https://linkedin.com/company/phuse)