

Single Day Event

Framework to select the “right” statistical programming language for standards development

Carl Herremans, MSD, Brussels, Belgium
Lino Jardim, MSD, Zurich, Switzerland

For presentation at PHUSE SDE
Brussels 2024



Disclaimer

This presentation does not necessarily reflect MSD's corporate vision. It reflects the authors' insights and diverse experiences working on analysis and reporting using various programming languages across the industry.

Why are we using standards?

Benefits of Standards

- High quality, validated, timeliness, cost-effective and reproducible deliveries.
- Economies of scale. The higher usage of standard code, the higher the return on investment.

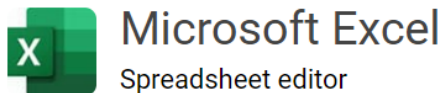
Costs of Standards

- Development/validation or qualification of code.
- Maintenance
- Documentation, communication and training.
- (Licensing)



Poll:

What statistical programming language would you use?



(*) All brands and products names are trademarks of their respective companies

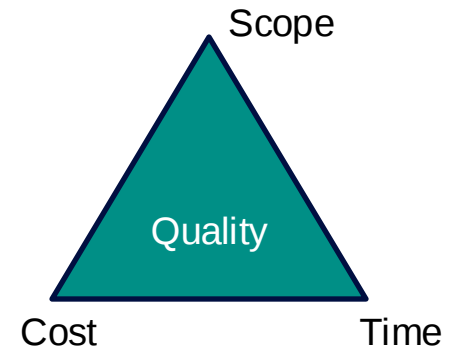
phuse What statistical programming language would you use?



People tend to use what they know

- *Knowledge and familiarity are good*
- *Learning curve is low*

But is this the “best” approach to choose a program language for a standard?



Is this language the “best” choice?

- To solve a one-off task, the choice is probably not needed.
- Dimensions to consider when developing a standard:

Process

- Is there a language agnostic SOP in place or is there a specific SOP for this language?

Accessibility

- Do the others in the team know this programming language?

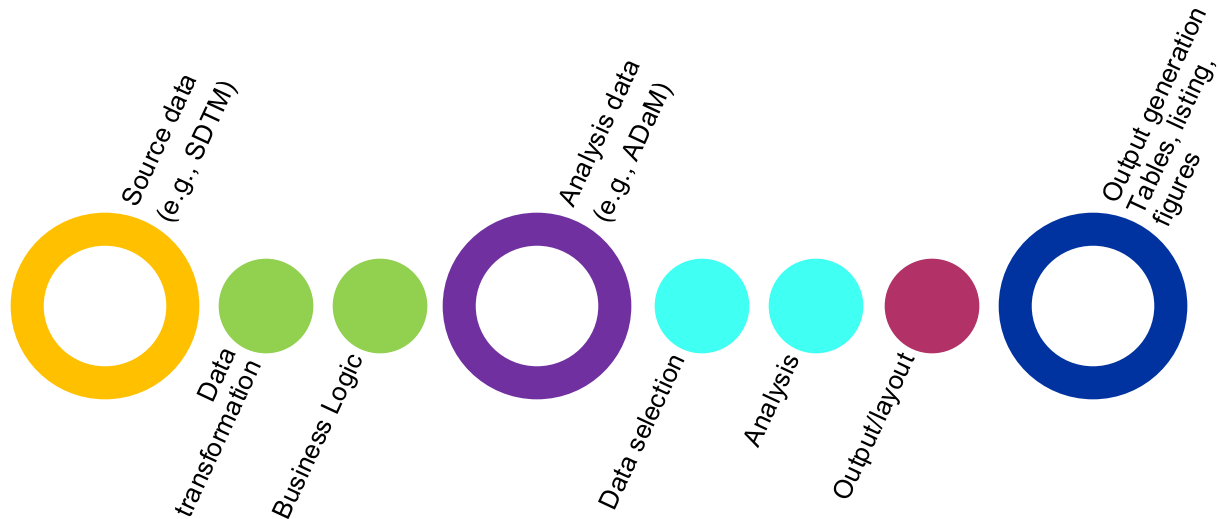
Reproducibility

- If a package from the internet is used, will it last?

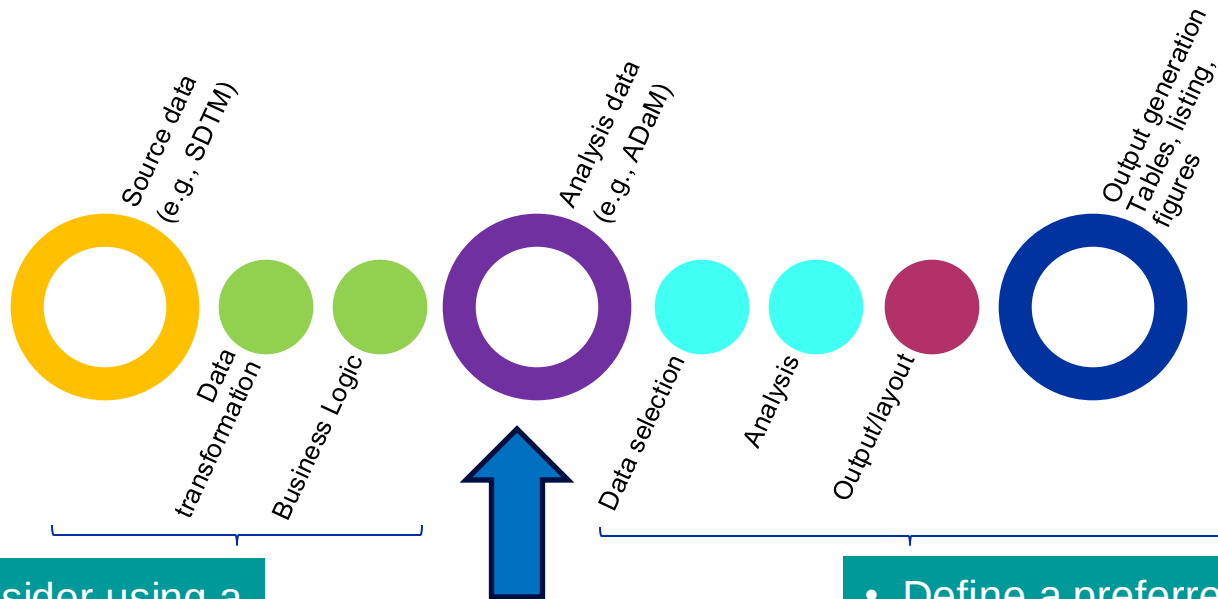
Compatibility

- Is it possible to line up the programs of different languages?

Typical Analysis and Reporting data flow



Typical Analysis and Reporting data flow: Recommendation



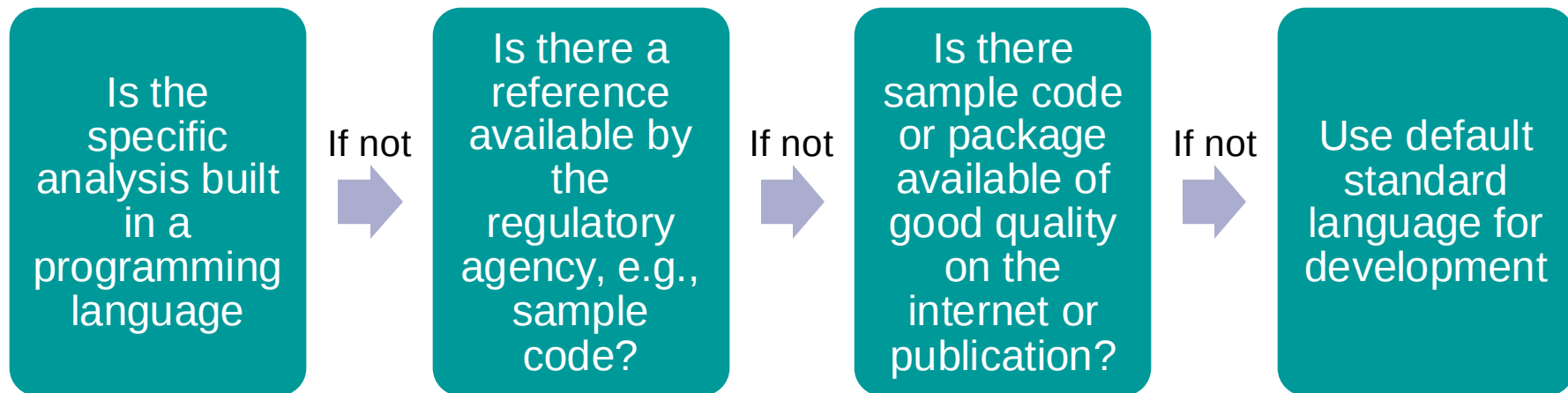
Consider using a single language

Define single data format accessible to all programming languages

- Define a preferred programming language, potentially by analysis cluster
- Consider different languages for analysis and reporting output generation

Is this language the “best” choice for this specific analysis?

- Use case: What analysis is exactly performed?
- Predefined criteria to consider before starting standards development.
Example:



Conclusion

Framework to select the “right” statistical programming language for standards development

Let true cost drive your analysis and reporting language selection

Multiple languages are a right solution for standardization analysis

Have a SOP & infrastructure in place to support multiple languages

Pick one programming language to convert your input data to analysis data

Define analysis dataset format suitable from all programming languages

Select a default language for developing a standard analysis but be open for multiple languages

Select analysis language based on an objective predefined set of criteria

Build the framework to select the “right” statistical programming language for standards development through

Language agnostic
management of
standards library

Intuitive to end
users

Streamline format
of trainings &
communications

Leveraging on best
practices across
programming
languages

Client centric:
A&R outputs are
the product not the
program code

We welcome any suggestions or observations:
Carl Herremans, carl.herremans@msd.com
Lino Jardim, lino.jardim@msd.com



**The Global Healthcare
Data Science Community**

Contact Channels

📞 UK +44 1843 609600
✉ office@phuse.global
🌐 phuse.global

Social Media

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