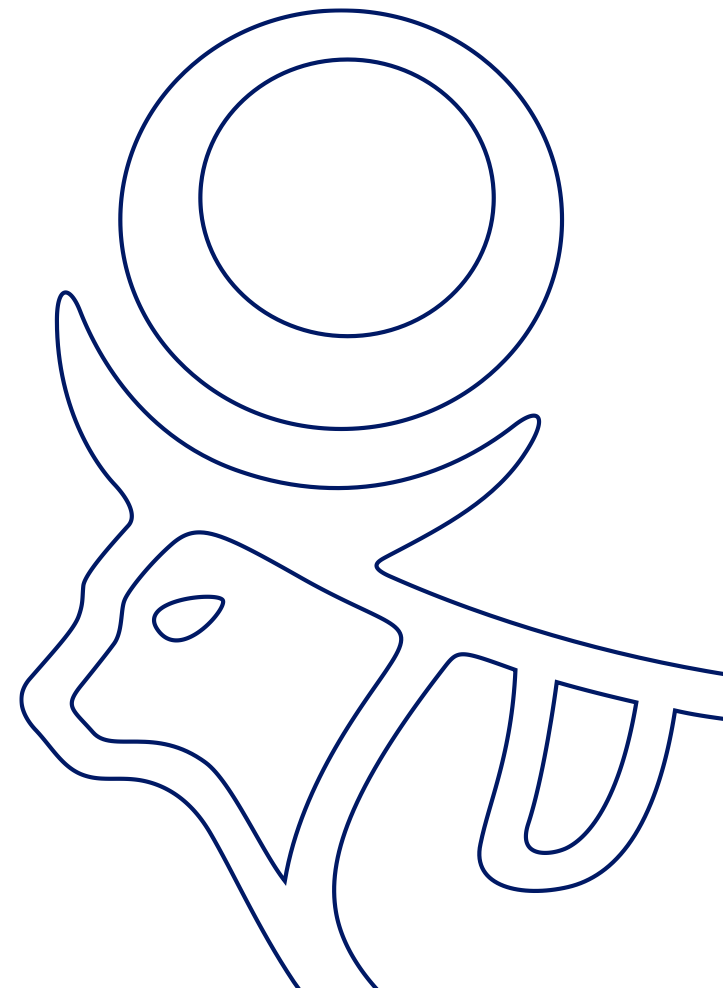


OS11

Aspirations of Language Agnosticism – Experience of Expanding the Programming Stack with Python

Presenter: Kaj Arai Hejlesen, Novo Nordisk A/S



Disclaimer

Views and opinions expressed are those of the speaker and not necessarily Novo Nordisk



"You can get it any color you want, as long as it is black"

Attr. Henry Ford



"You can use any programming language you want, as long as it is SAS"

The pharma industry

R usage in the industry

Novo Nordisk: Journey to an R-based FDA submission

2023-08-11

Johnson & Johnson's Open Source Journey in R

2025-04-17

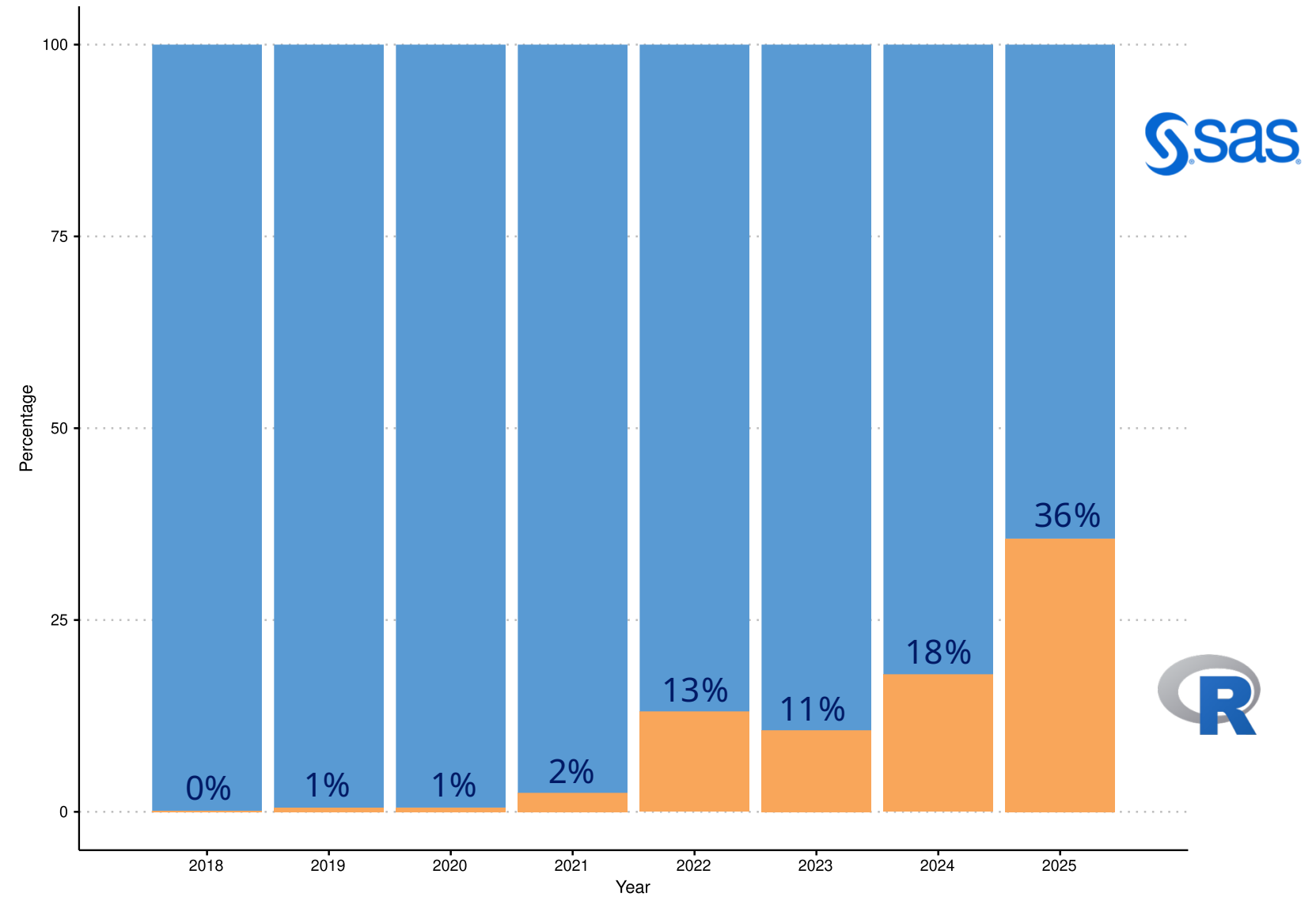
Roche's end-to-end R submission for a new drug application

2024-08-01



Source: posit.co

R and SAS usage in Novo Nordisk



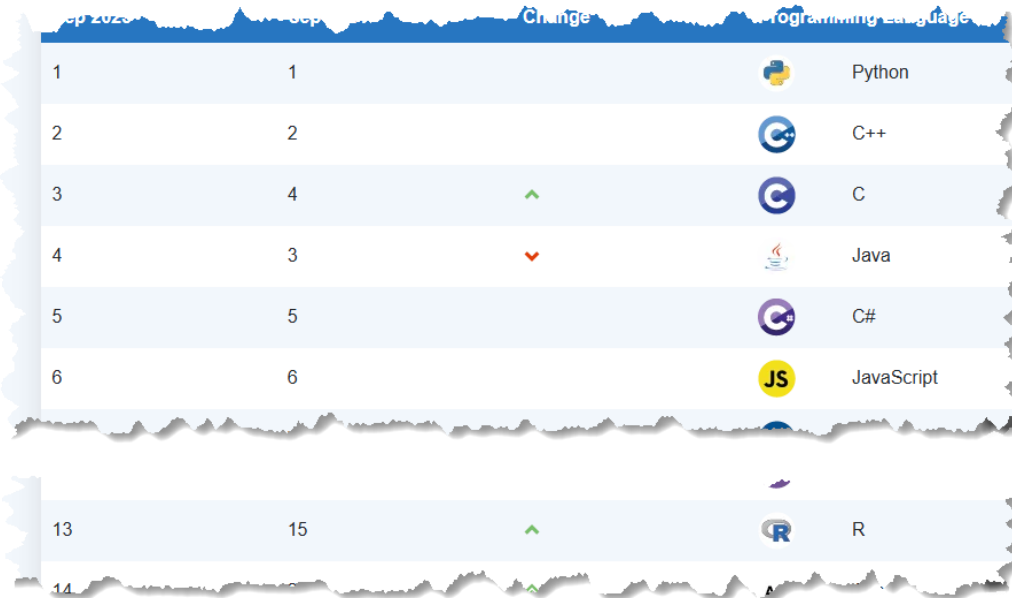
Why do we need more?

Let developers choose their preferred tool for the task at hand

More python developers than R or SAS – easier to recruit

New data modalities: images + omics

- GCP requirement
- R/SAS do not offer all features



The image shows a screenshot of the TIOBE Programming Language Index website. The table lists programming languages by their popularity index. Python is at the top with a rank of 1. R is at rank 13. The table includes columns for rank, change, and language name with its icon.

Rank	Change	Programming Language
1		Python
2		C++
3	▲	C
4	▼	Java
5		C#
6		JavaScript
13	▲	R

Source: www.tiobe.com

Purpose of the experiment

1. Create a GCP-validated delivery to gain experience and open the floodgates

2. delivery NOT in scope for submitting programs to authorities



Choice of delivery

PERIODIC SAFETY UPDATE REPORT (PSUR)



ADaM



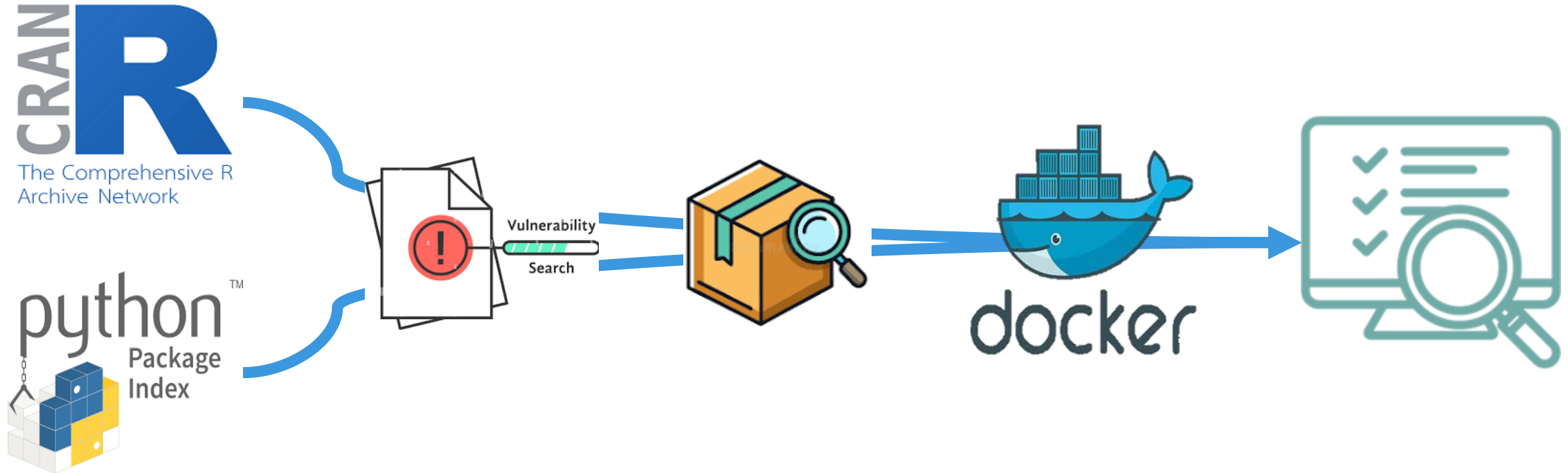
TFL



Parallel review



GCP validation of packages and environment



We have our environment – how do we start?



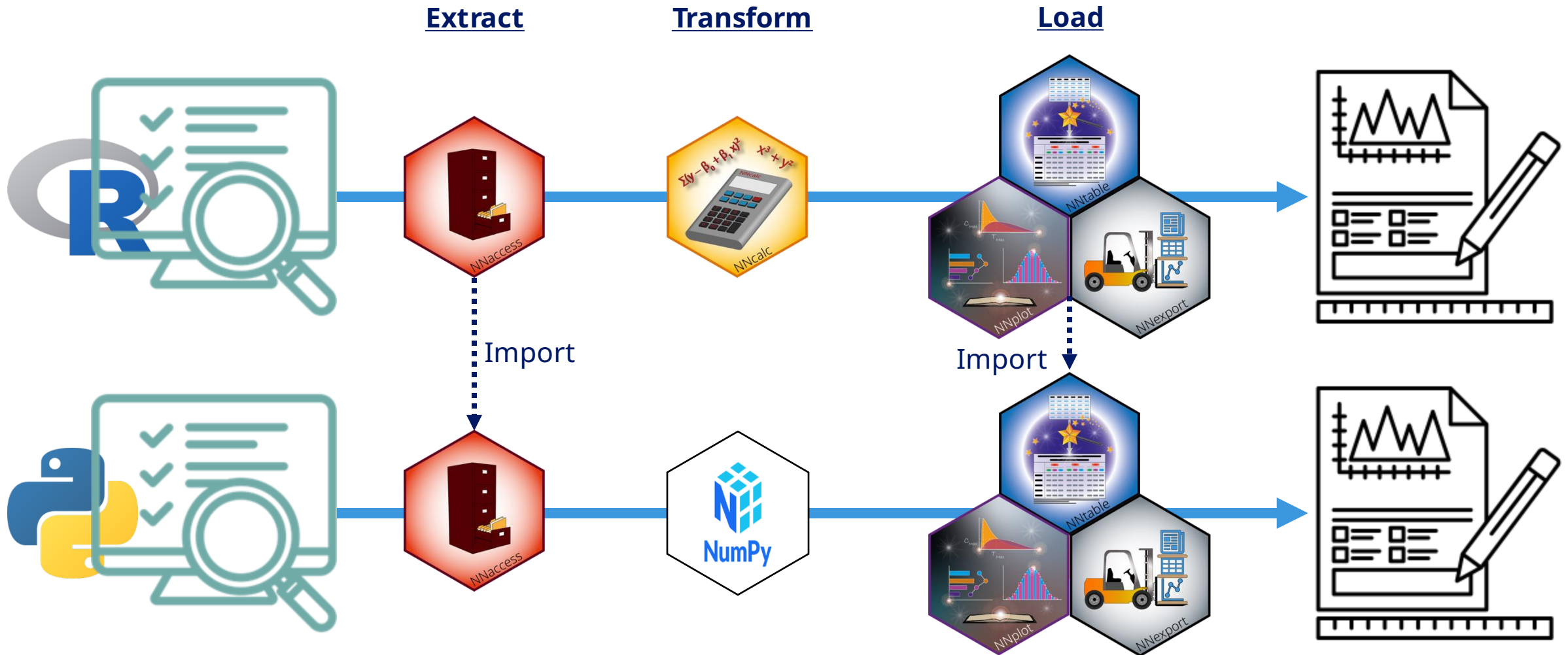
Project overview:

- Have 2 FTEs at 50% allocated for 2.5 months
- Make 4 TFL programs
- Generate 19 outputs

Output requirements:

- Must apply usual styling which our stakeholders are expecting
- Outputs must be a specific file format

From compute environment to TFL report



Rpy2



```
# R package imports
```

```
from rpy2.robjects.packages import importr
```

```
NNaccess = importr('NNaccess')
```

```
db = NNaccess.nnaccess()
```

```
✓ print(db.names) ...
```

[1] "adam"	"adam_tmp"	"xprt"	"adam_map"
[5] "sdtn"	"sdtn_map"	"adm"	"external_data"
[9] "metadata"	"output_datasets"	"data"	"gcmd"
[13] "build_data"	"exportOutput"	"paths"	

The Execution



<i>str</i>	<i>int</i>	<i>float</i>	<i>date</i>	<i>datetime</i>	<i>time</i>
"apple"	1	1.849	13SEP2024	14SEP24:08:56:05	13:42:37



rpy2 – haven::read_sas()

Through internal R package



<i>object</i>	<i>float</i>	<i>float</i>	<i>float</i>	<i>datetime</i>	<i>float</i>
apple	1.	1.849	19979	2024-09-14 08:56:05	49357



The Execution

Default behaviour of python packages might differ:



`numpy.std()` defaults to using population variance

$$\sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$$



`proc means` defaults to using sample variance

$$\sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}$$

Project outcome



Project overview:

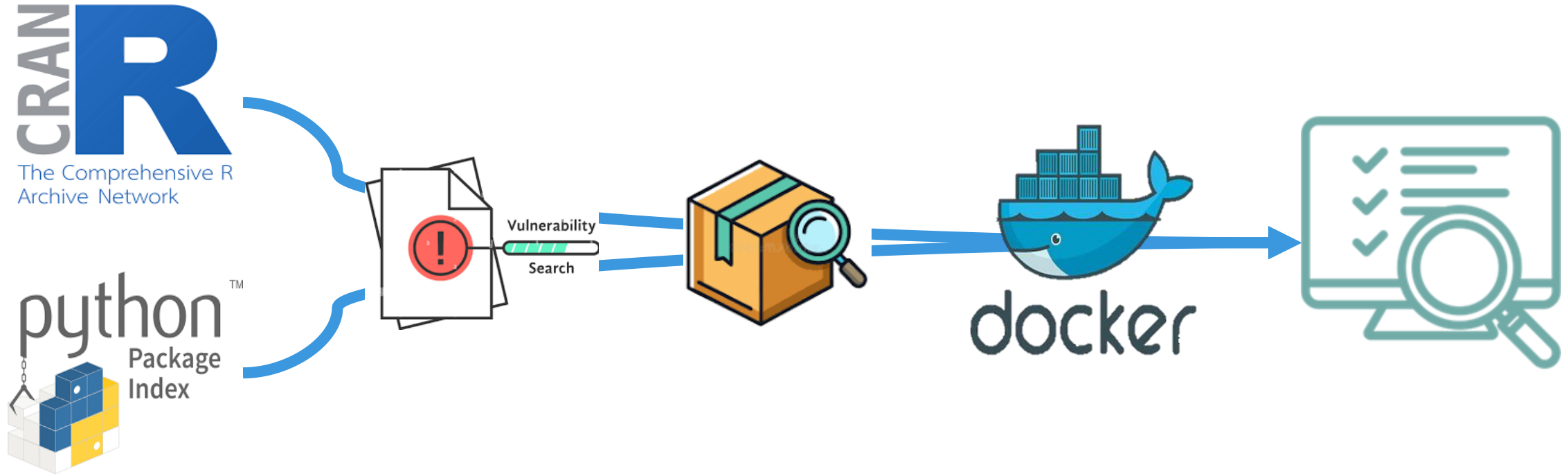
- Have 2 FTEs at 50% allocated for 2.5 3.5 months
- Make 4 TBLT programs
- Generate 19 73 outputs

Output requirements:

- Must apply usual styling which our stakeholders are expecting
- Outputs must be a specific file format



Validation is just the start!



Dedicated support for each language

- rpy2 is not a magic one-fix tool, beware!
- Be mindful of default behaviours of statistical tools. Test!
- maintenance needed!

rpy2

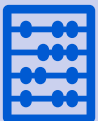
Standards, standards, standards



Don't use language dependant **data formats**: sas7bdat, Rds

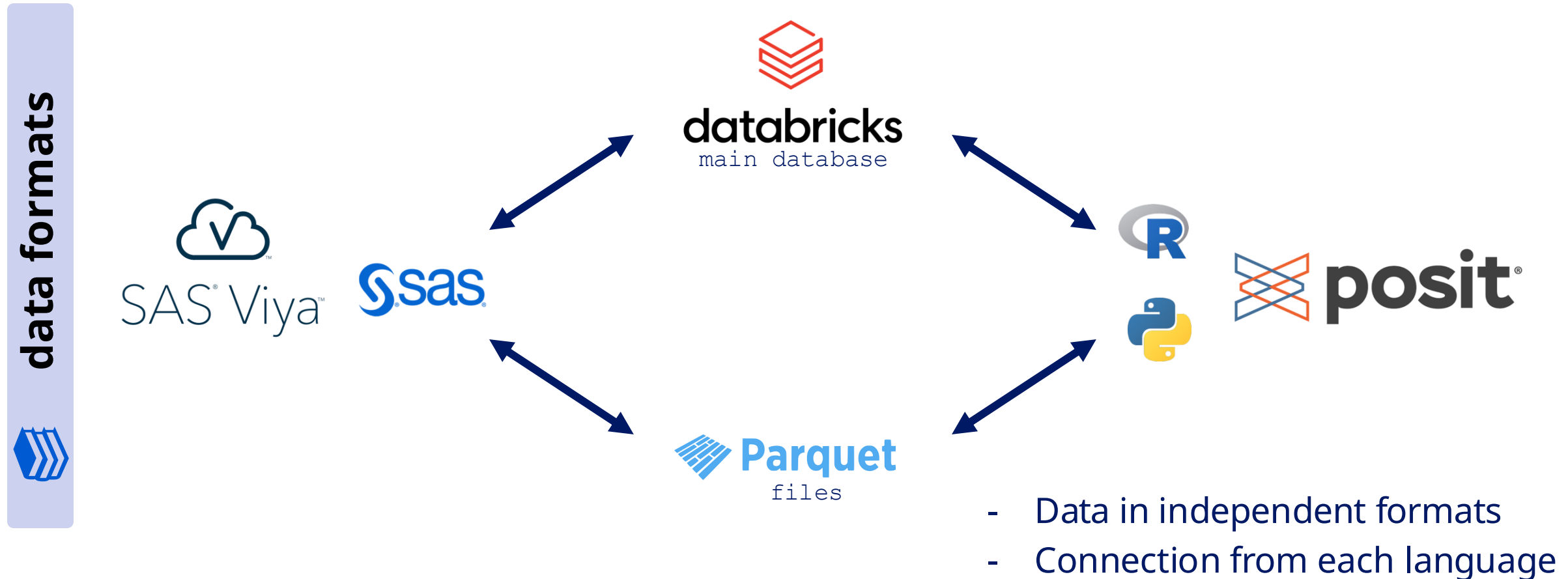


Make language-independent **specifications**

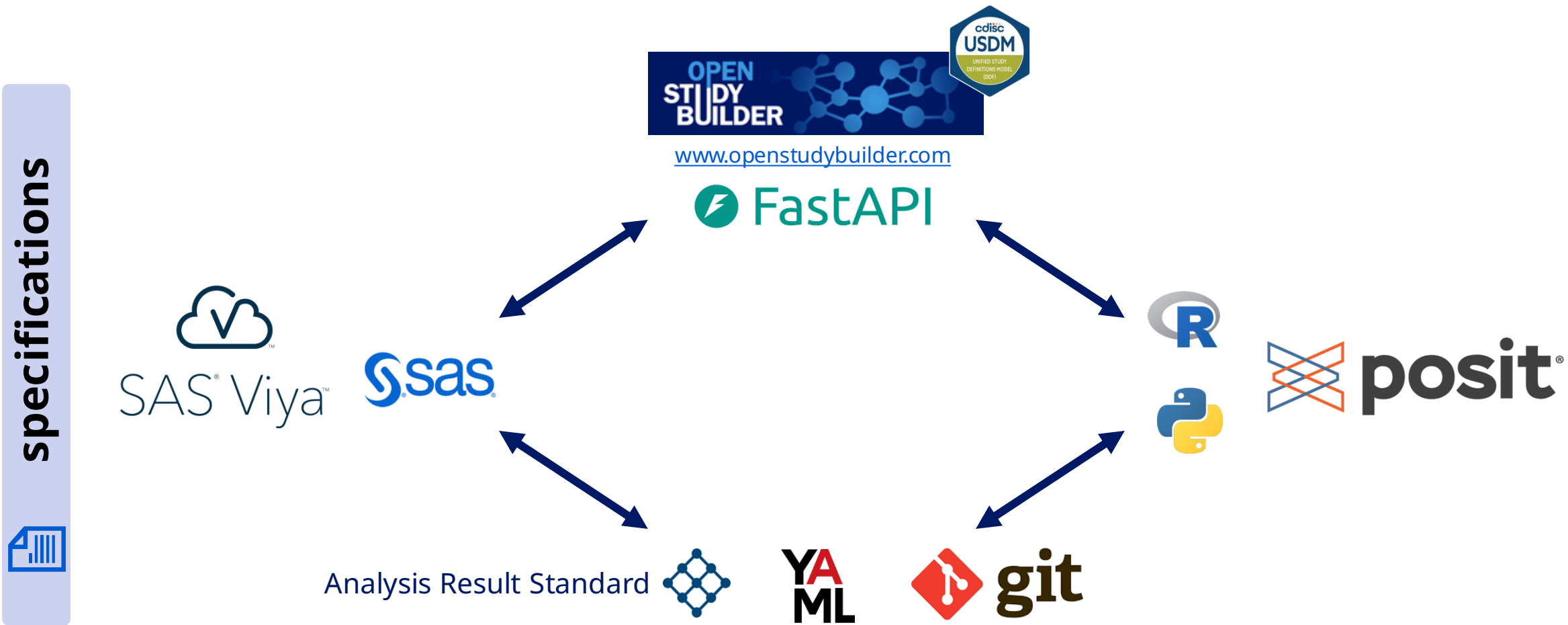


Use **tools** supporting multiple languages

Standards, standards, standards



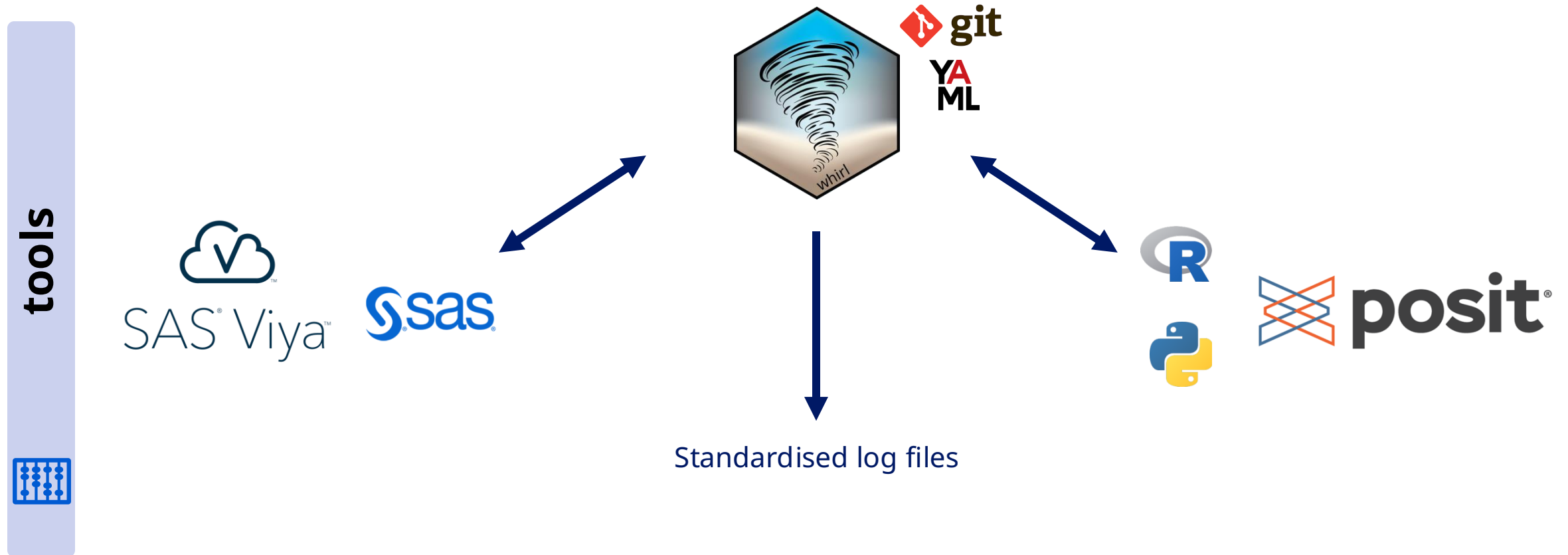
Standards, standards, standards

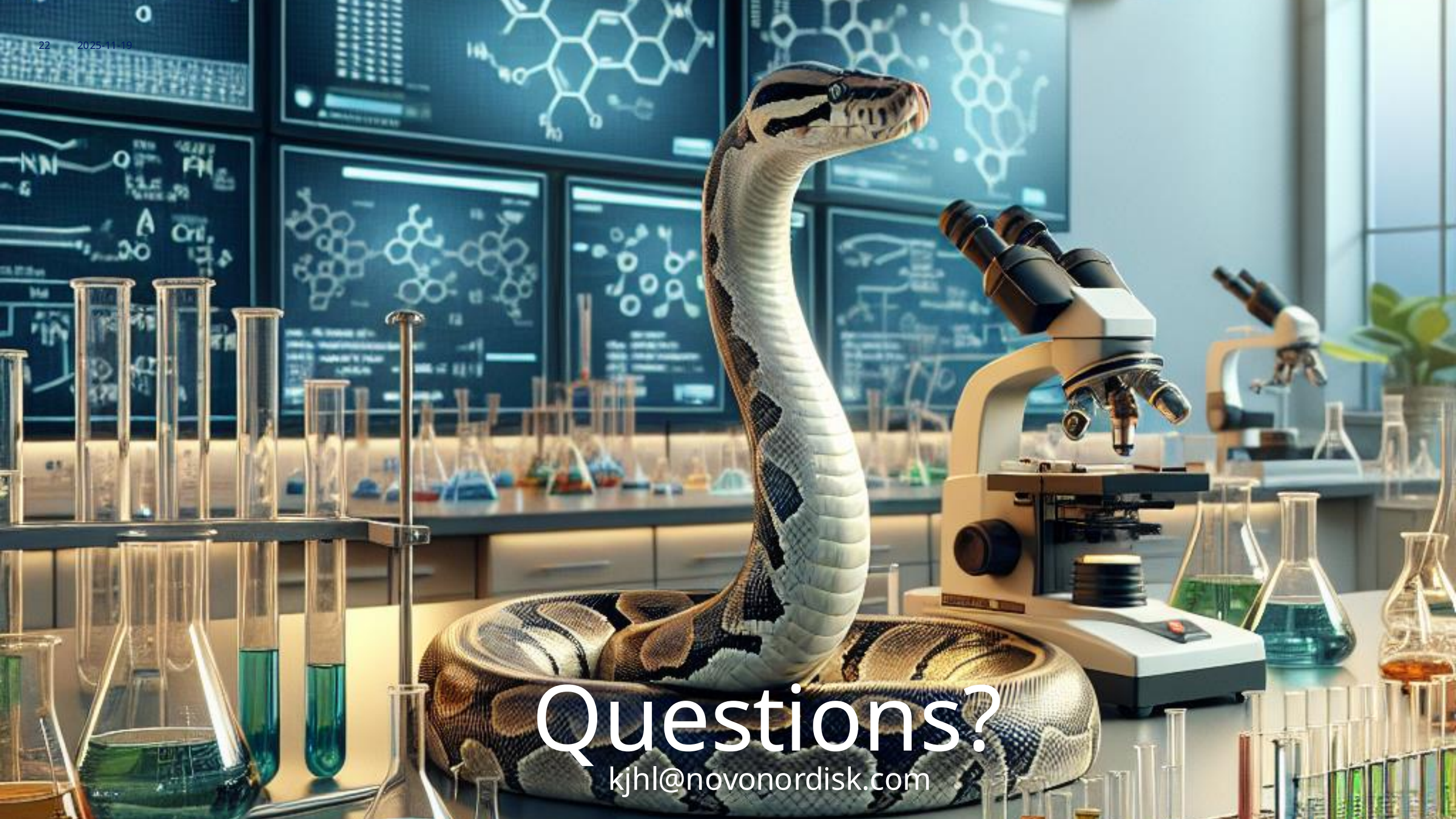


- Centralised repo essential

Standards, standards, standards

Case: program execution





Questions?

kjhl@novonordisk.com