

BBMRI.nl

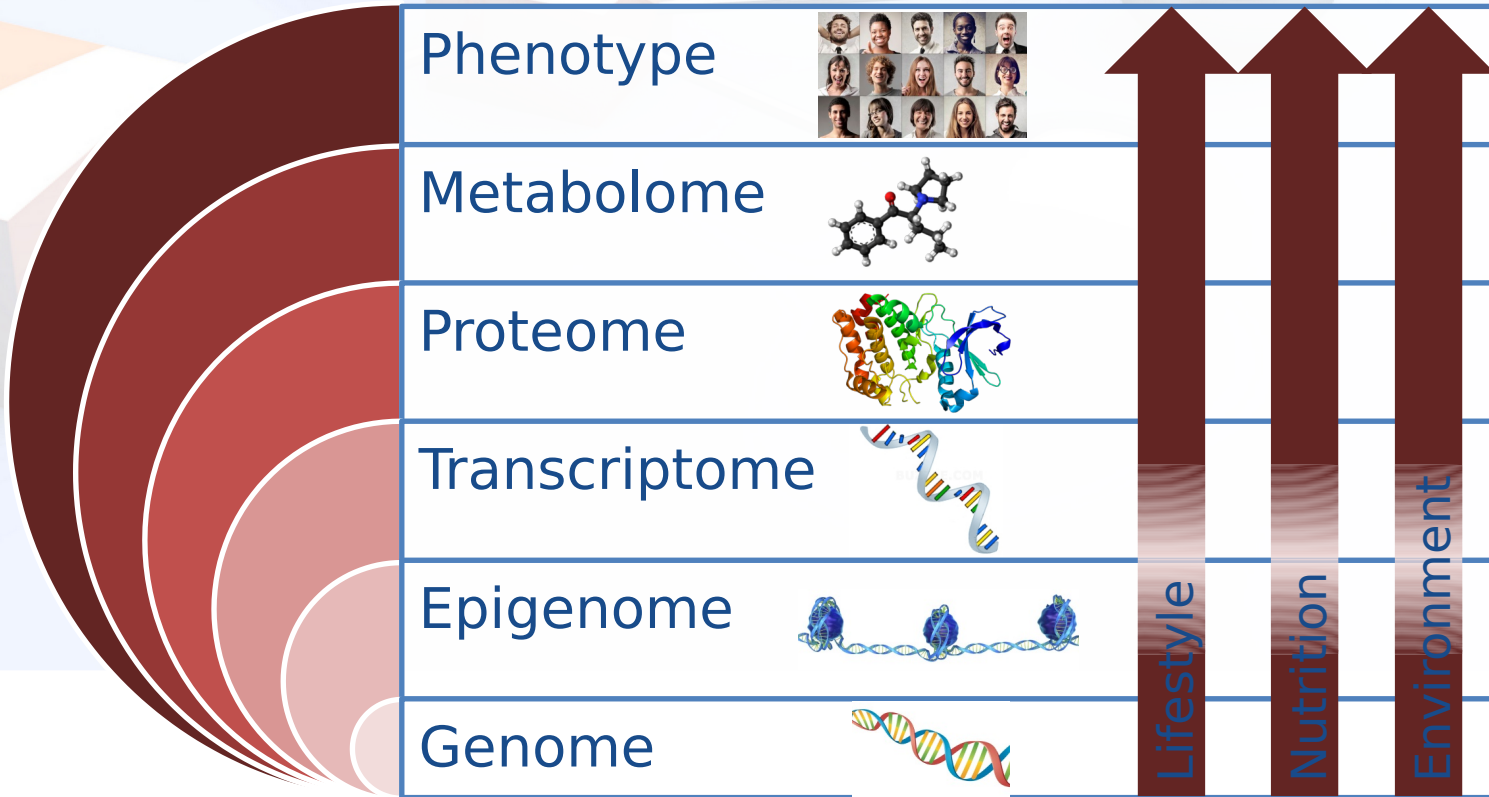
Biobanking and
BioMolecular resources
Research Infrastructure
The Netherlands

Data infrastructure of the BBMRI-omics project

Leon Mei, Leiden University Medical Center

AMC, 14 October 2019

BBMRI-Omics



BBMRI-Omics: molecular data (2010 -)

N~100,000
GWAS

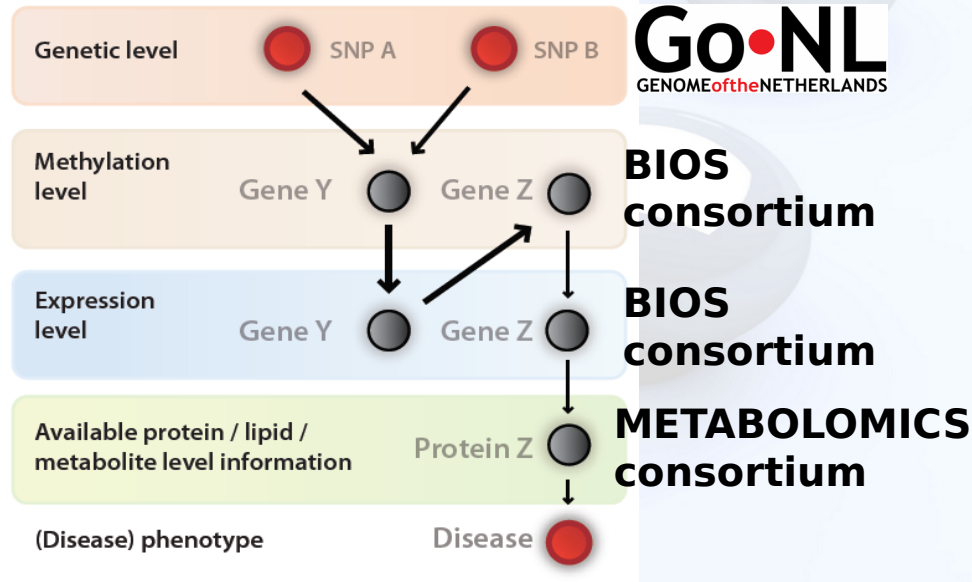
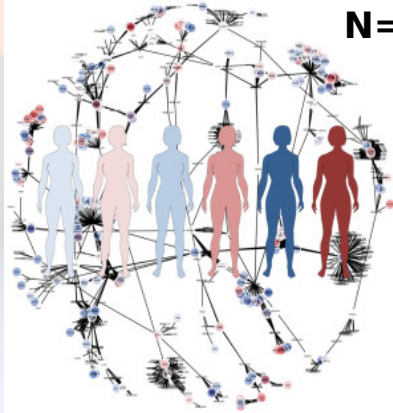
N=770 **Go.NL**

N~4,000

N~4,000

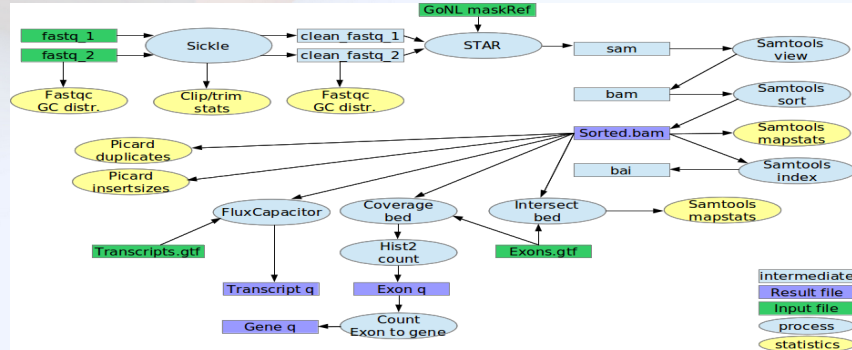
N~50,000

N>250,000



Building a central database for the research community

- To store raw and processed data
- Reference data for case-control studies and imputation
- Linking to clinical phenotypes



BIOS-VM

```
hmei@bios-vm: ~
hmei@Leon-LUMC:~$ ssh bios-vm.bbmrirp3-lumc.surf-hosted.nl
Welcome to Ubuntu 16.04.6 LTS (GNU/Linux 4.4.0-154-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:    https://landscape.canonical.com
 * Support:        https://ubuntu.com/support

56 packages can be updated.
0 updates are security updates.

New release '18.04.2 LTS' available.
Run 'do-release-upgrade' to upgrade to it.

*** System restart required ***

      _ _ _
     / / /
    / / /
   / / /
  / / /
 / / /
/_/_/_

Welcome to the cloud VM for BBMRI

 * Documentation: http://www.bbmrirp3-lumc.nl
Last login: Sun Oct 13 22:07:57 2018
hmei@bios-vm:~$ ls /virdir/Backup/R/
Genotypes      GWAS_Imputation
gEUVADIS       HRC_Imputation
gonl-snp-indels-release-5  HRCv1.1_Imputation
gonl-snv-release-5.4      IlluminaHumanMethylation450K
                                     Phenotypes
                                     README
                                     reference
                                     RNASeq
                                     RP4

hmei@bios-vm:~$
```

RStudio - Mozilla Firefox

bios-vm.bbmrirp3-lumc.surf-hosted.nl/rstudio/

File Edit Code View Plots Session Build Debug Profile Tools Help

Environment History Connections

Global Environment

Environment is empty

Files Plots Packages Help Viewer

New Folder Upload Delete Rename More

virdir Backup RP3_data

Name Size Modified

Genotypes

gEUVADIS

gonl-snp-indels-release-5

gonl-snv-release-5.4

GWAS_ImputationGoNLV5

HRC_Imputation

HRCv1.1_Imputation

IlluminaHumanMethylation450K

Phenotypes

README 8 KB Mar 13, 2017, 4:28 PM

reference

RNASeq

RP4

Demos in package 'base':

error.catching More examples on catching and handling errors

is.things Explore some properties of R objects and is.FOO()

recursion Using recursion for adaptive integration

scoping An illustration of lexical scoping.

Demos in package 'graphics':

Hershey Tables of the characters in the Hershey vector font

Japanese Tables of the Japanese characters in the Hershey vector font

graphics A show of some of R's graphics capabilities

image The image-like graphics builtins of R

png Extended png() examples

Console Terminal

/mnt/virdir/Backup/hmei/testR/

Copyright (C) 2017 The R Foundation for Statistical Computing

Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.

You are welcome to redistribute it under certain conditions.

Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.

Type 'contributors()' for more information and

'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or

'help.start()' for an HTML browser interface to help.

Type 'q()' to quit R.

> plot(1)

>

@SURFsara

HPC cloud

~50 users

64core

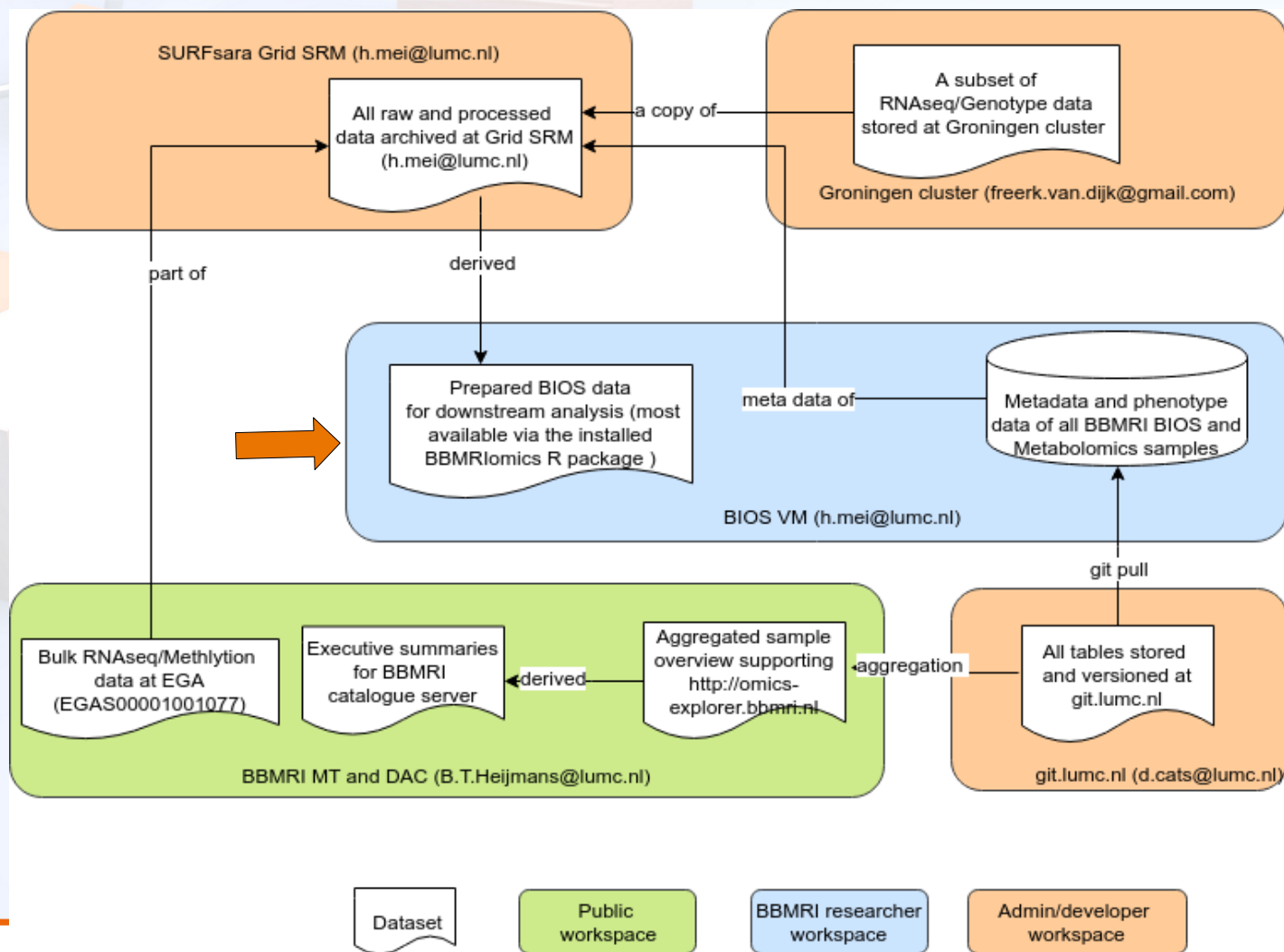
250G mem

5TB

CoC

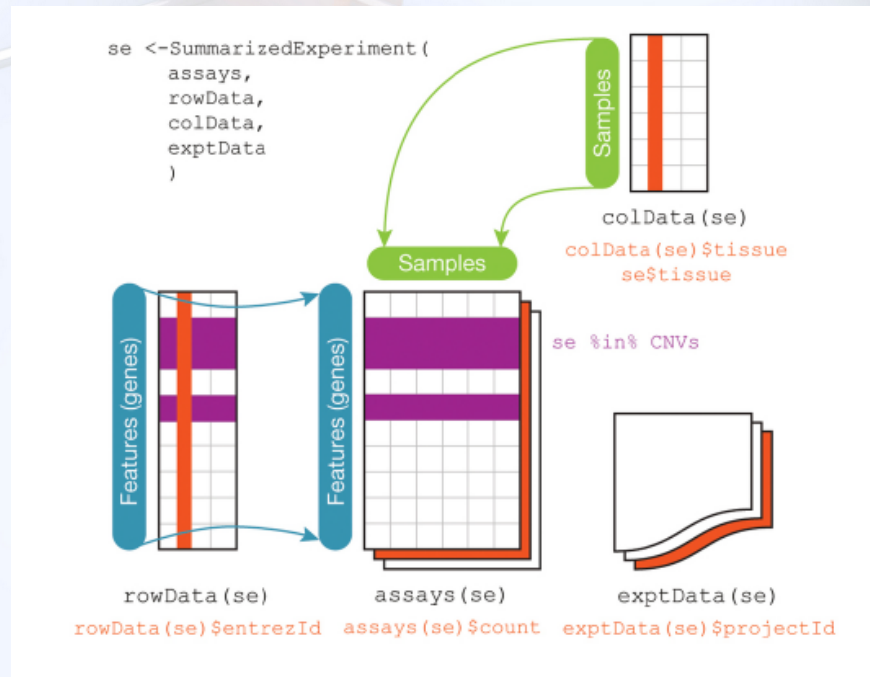
PKI auth

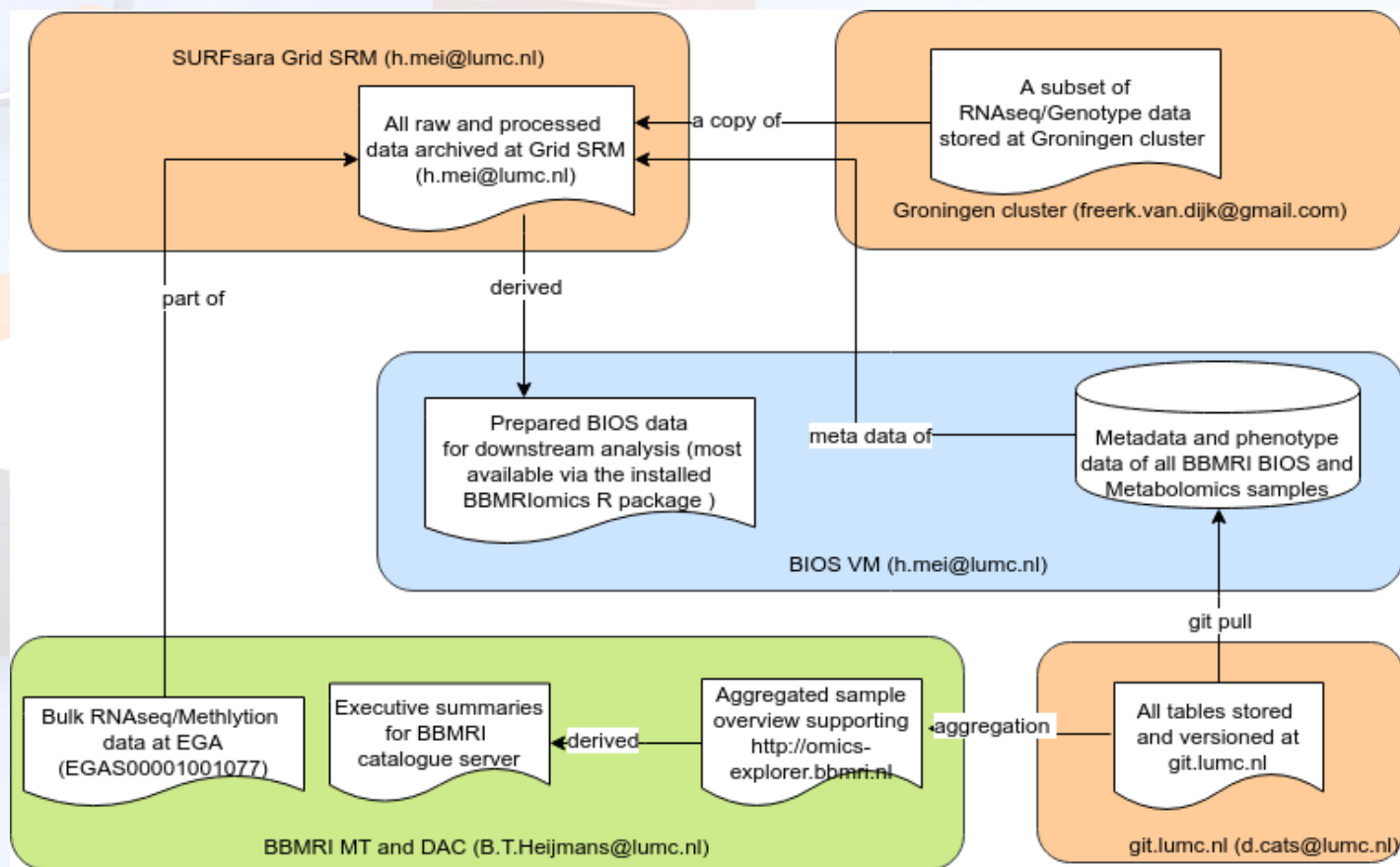
IP whitelisting



Ready to use data

BBMRIomics: An R interface to BIOS data



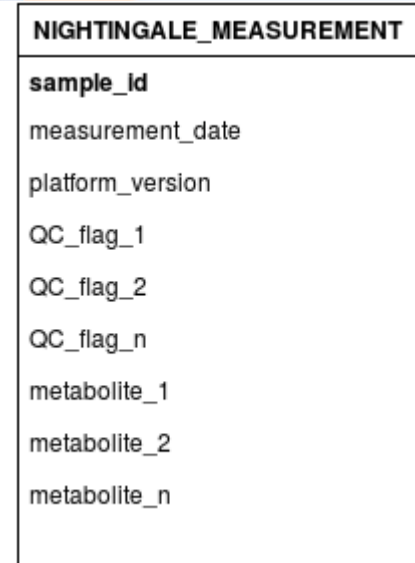


Dataset

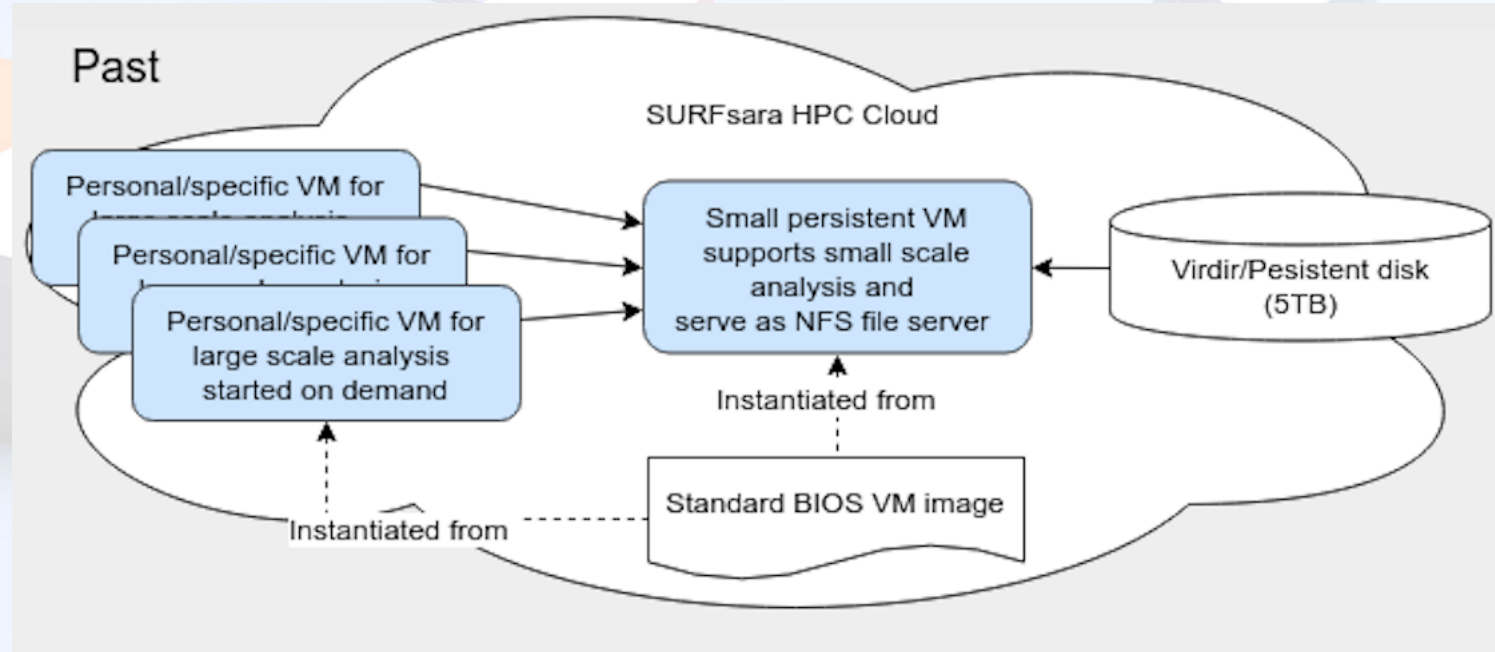
Public
workspace

BBMRI researcher
workspace

Admin/developer
workspace

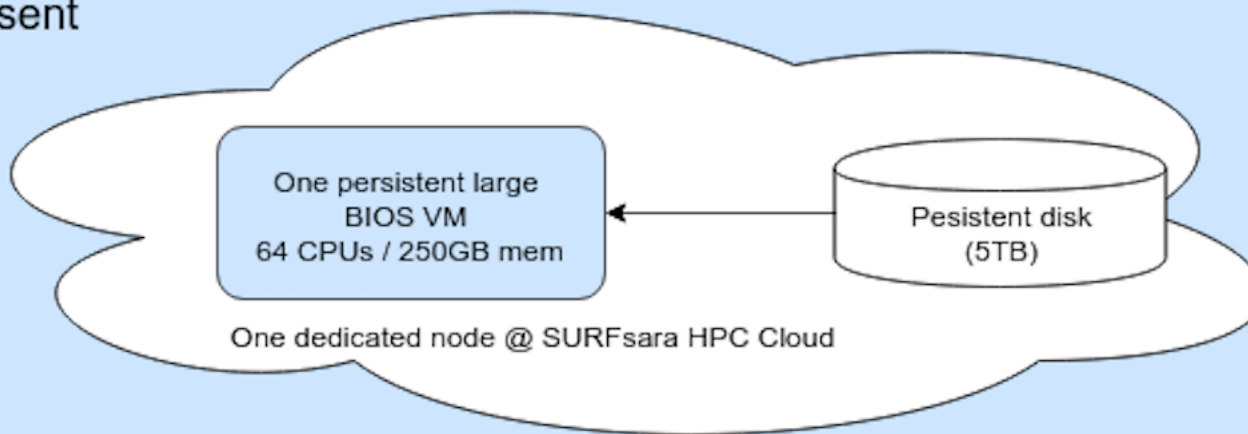


BIOS VM Past



BIOS VM Present

Present



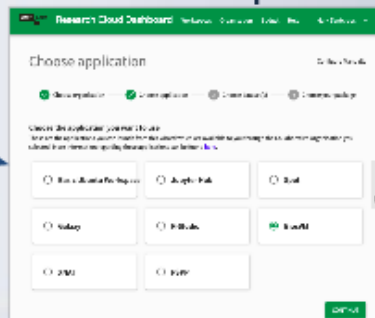
BBMRIomics workspace

Homepage



Institutional Login

Choose workspace

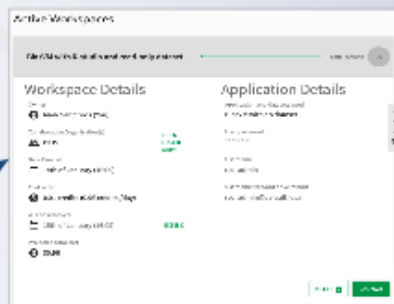


Rstudio login

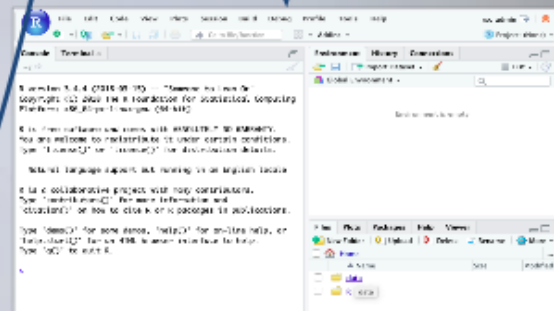


Data location

Resource dimensions

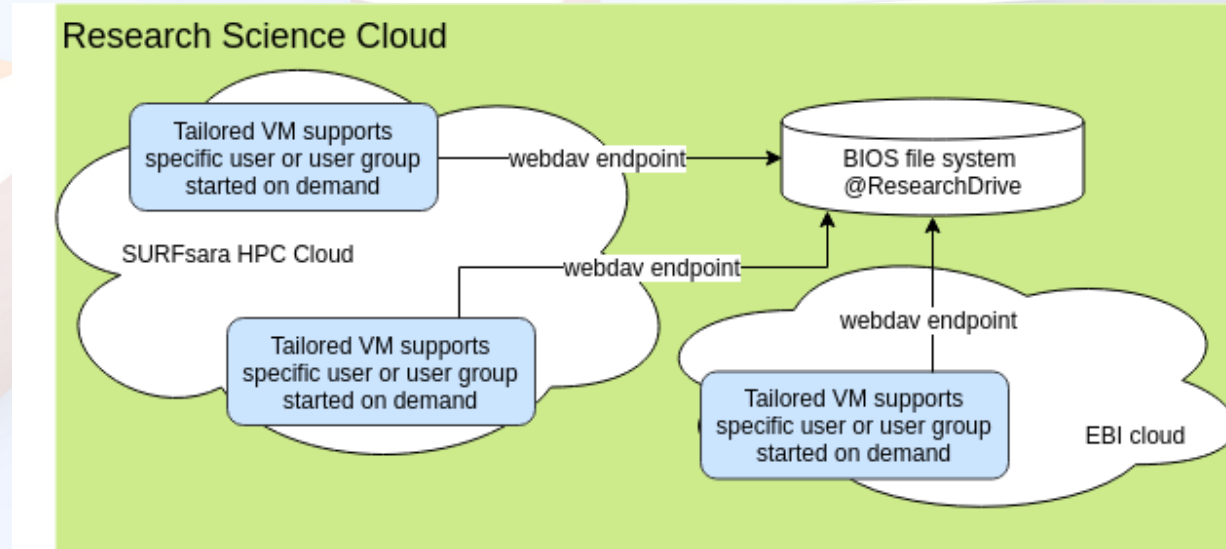


Launch workspace

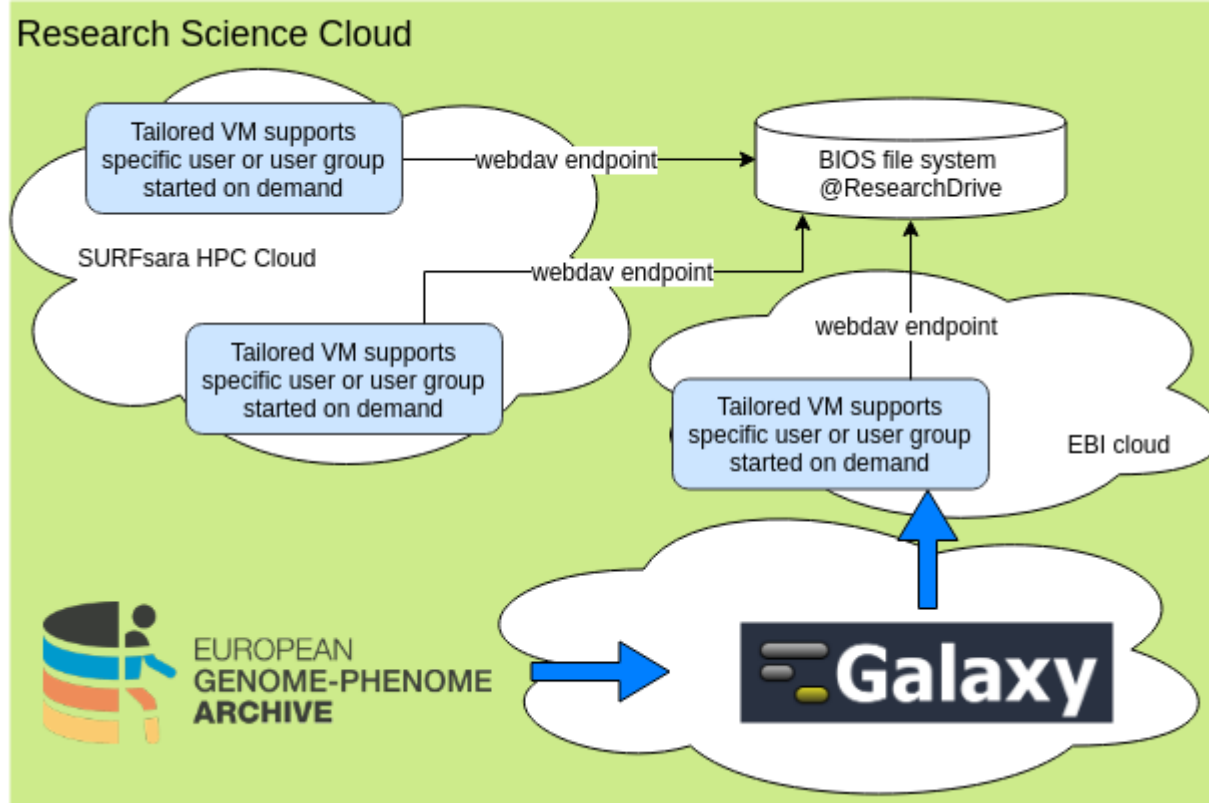


Ready for analysis!

BIOS VM Future



BIOS VM Future (VRE?)



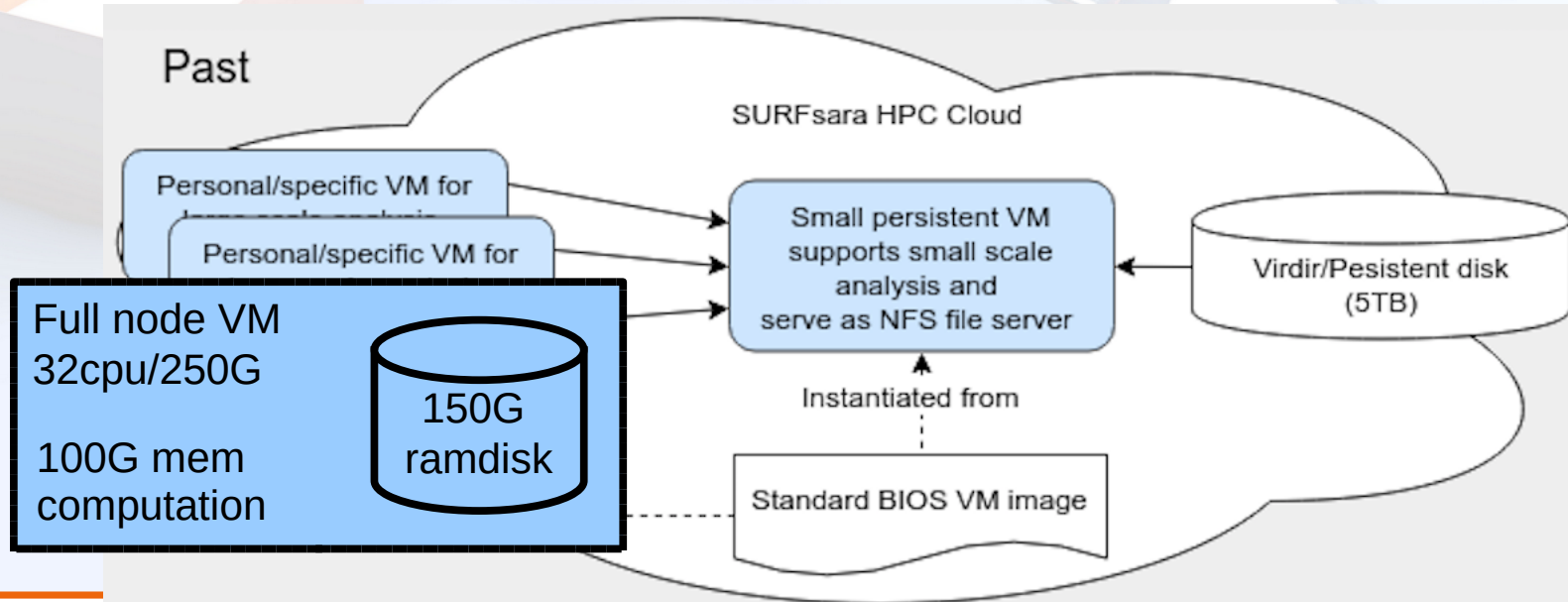
Lessons learned

Life Science Grid

- Programming for Grid is not easy
 - close collaboration with Grid experts (Jan Bot)
- Debug on Grid is not easy
 - create sufficient logging, progress flags, environment variables, library paths, etc.
- Work on Grid can be sometimes frustrating
 - Team or pair programming
- Running successfully on UI machine can not guarantee things will run on Grid clusters
- Data staging, e.g., use a local copy of frequently accessed files (reference genomes).

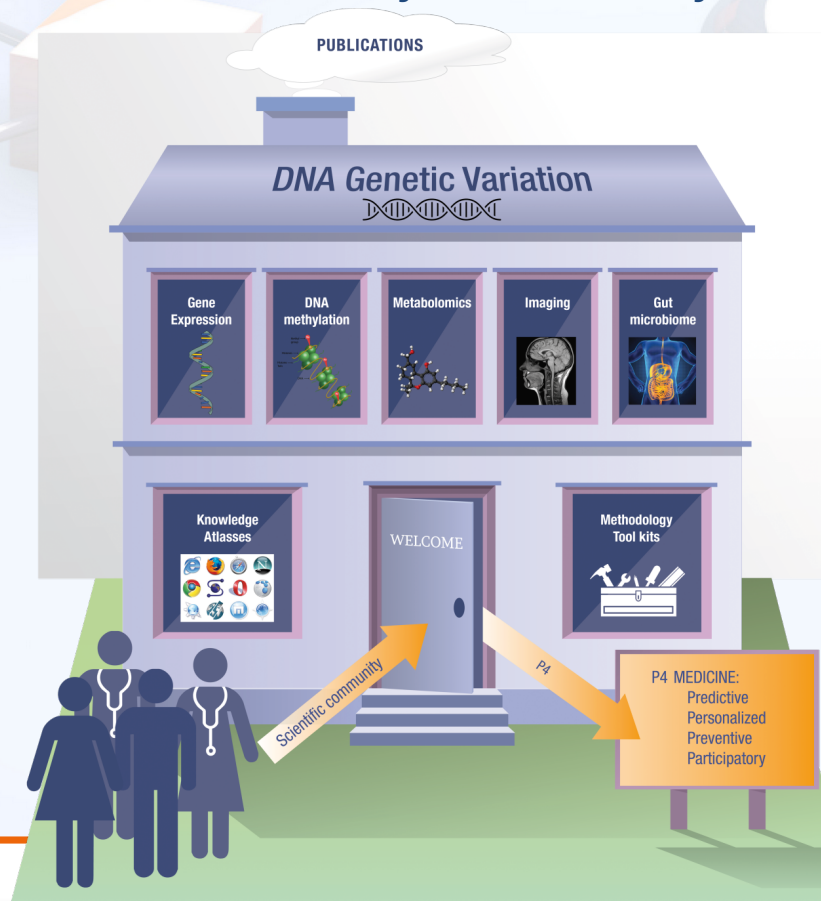
Cloud

- eQTL tests SNP and gene expression association
- The test tool requires a lot of I/O during the execution
- Suggestion from Mathijs Kattenberg (>50 times speed up)



BBMRI-OMICS:

Warehouse of data shared and used by the community





BBMRI-OMICS

For many, by many*

<https://www.bbmri.nl/>

**RDNL - Dutch Data
Prize 2018!**



SURF SARA

*100+ researchers working on 25 Cohorts, BIOS, GoNL, Metabolomics
Rainbow together building BBMRI-omics.