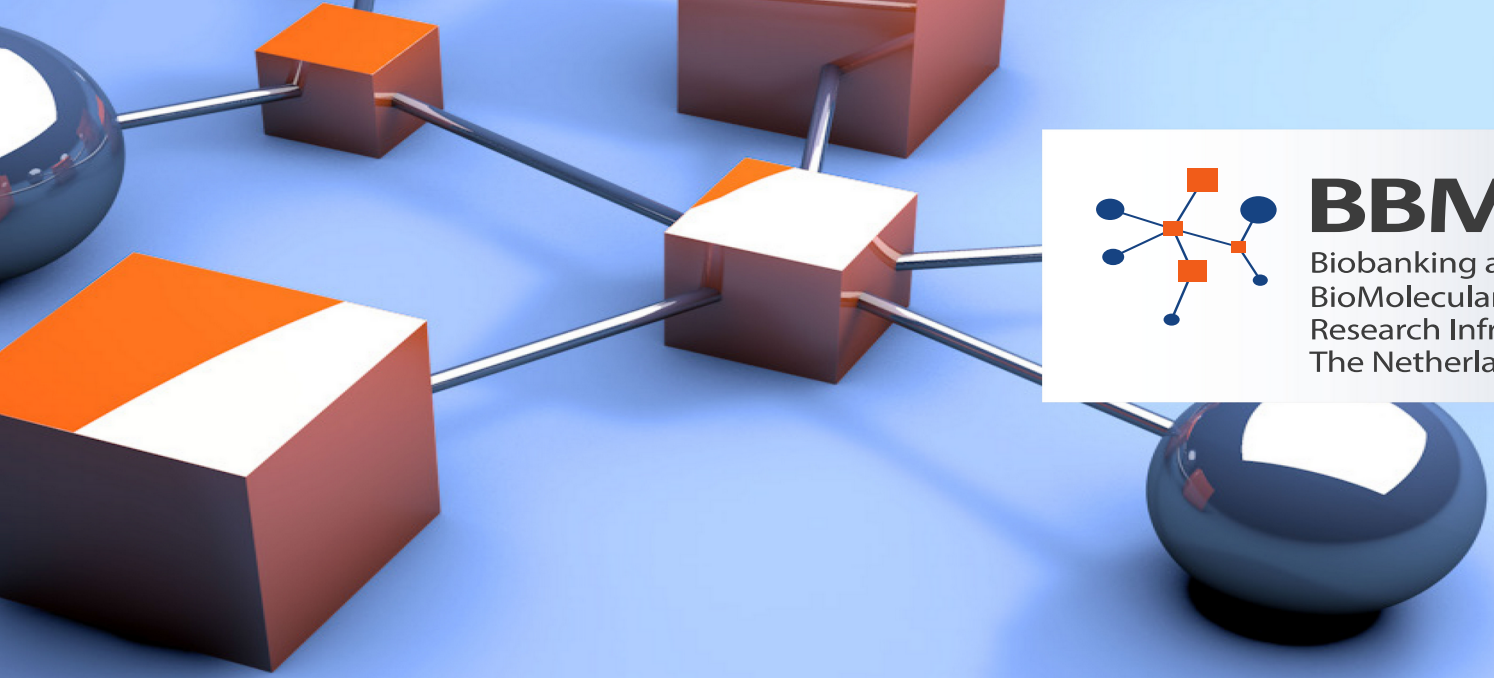


BBMRI.nl
Biobanking and
BioMolecular resources
Research Infrastructure
The Netherlands

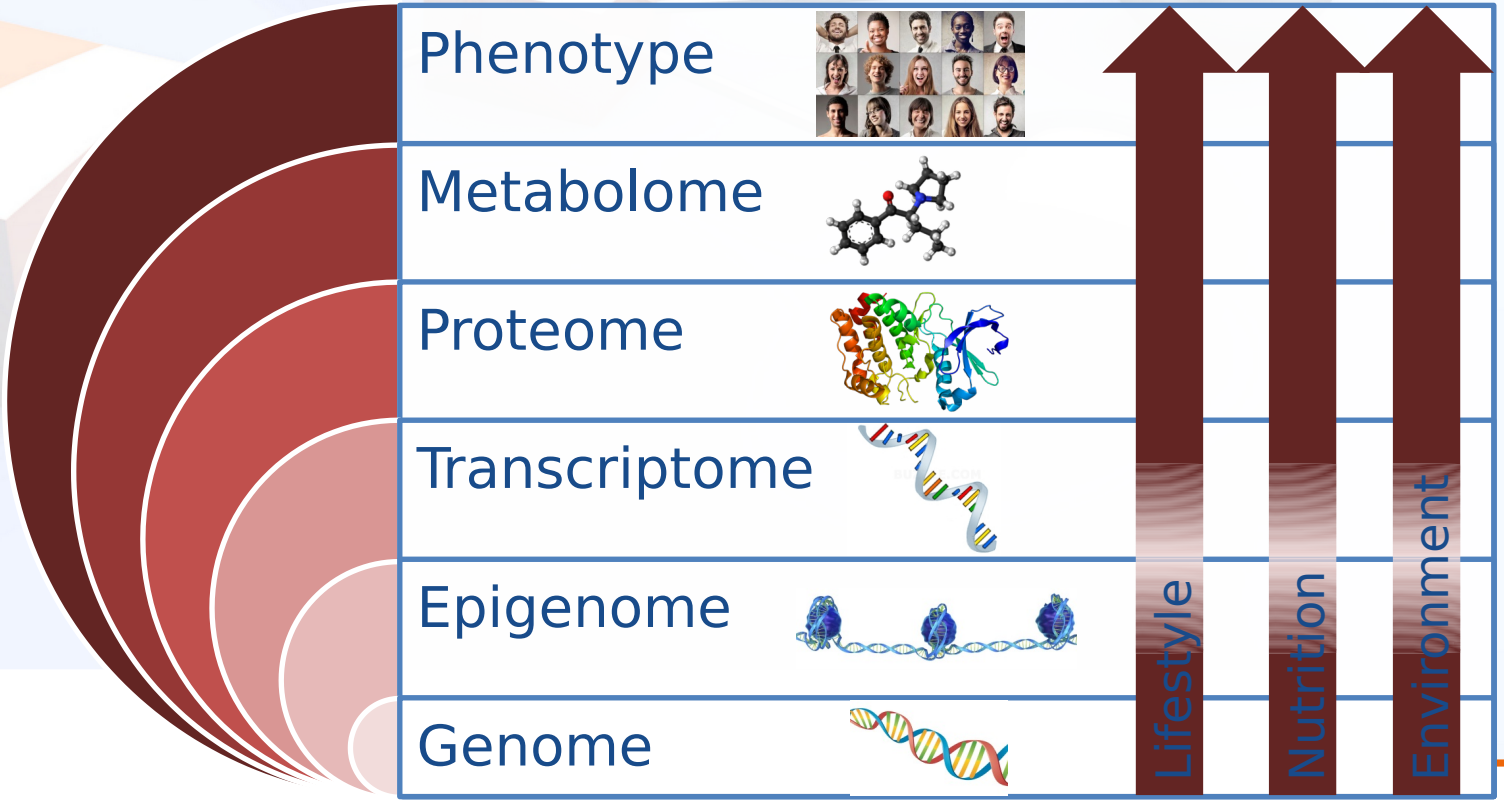
Data infrastructure of the BBMRI-omics project

Leon Mei, Leiden University Medical Center

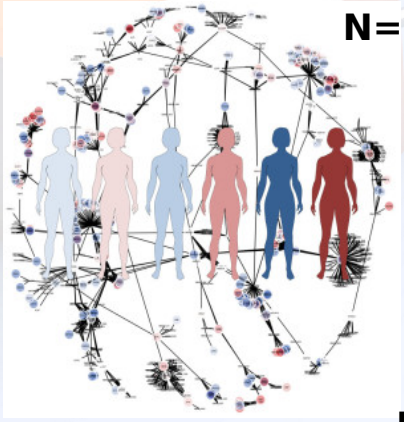
AMC, 14 October 2019

BBMRI.nl
Biobanking and BioMolecular resources Research Infrastructure The Netherlands

BBMRI-Omics



BBMRI-Omics: molecular data (2010 -)



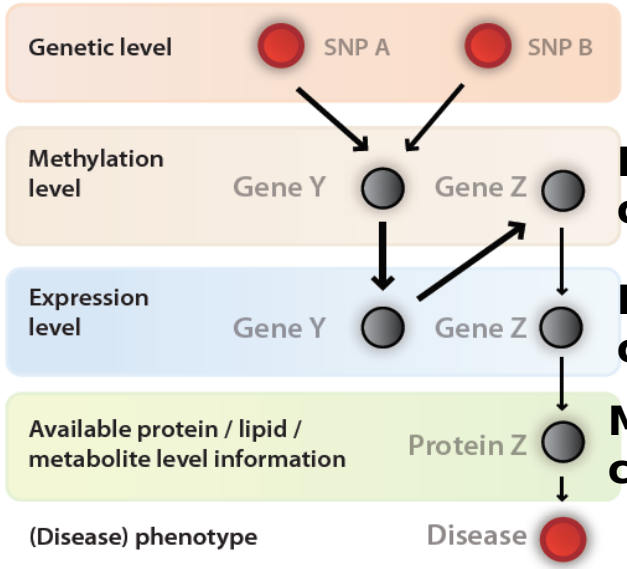
N~100,000 GWAS
N=770 Go.NL

N~4,000

N~4,000

N~50,000

N>250,000



Go•NL GENOMEoftheNETHERLANDS

BIOS consortium

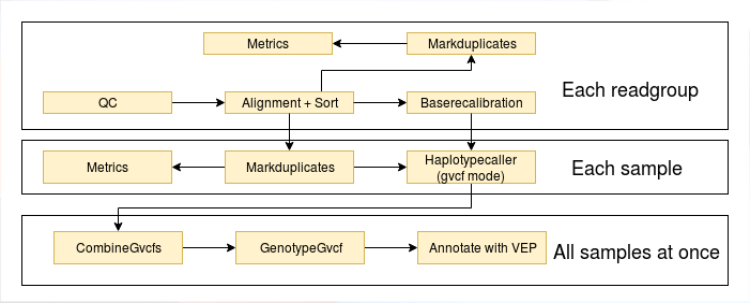
BIOS consortium

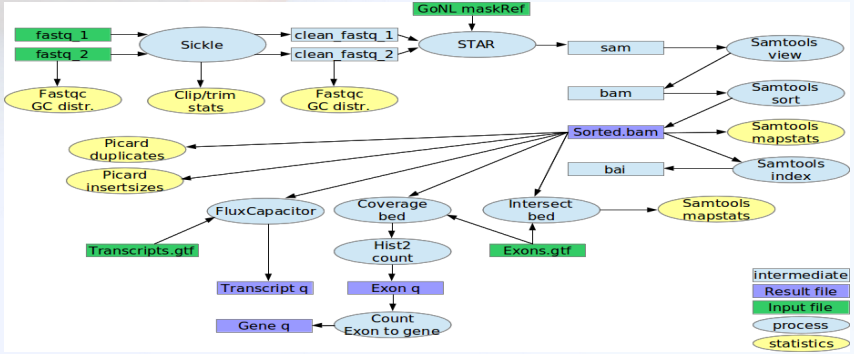
METABOLOMICS consortium

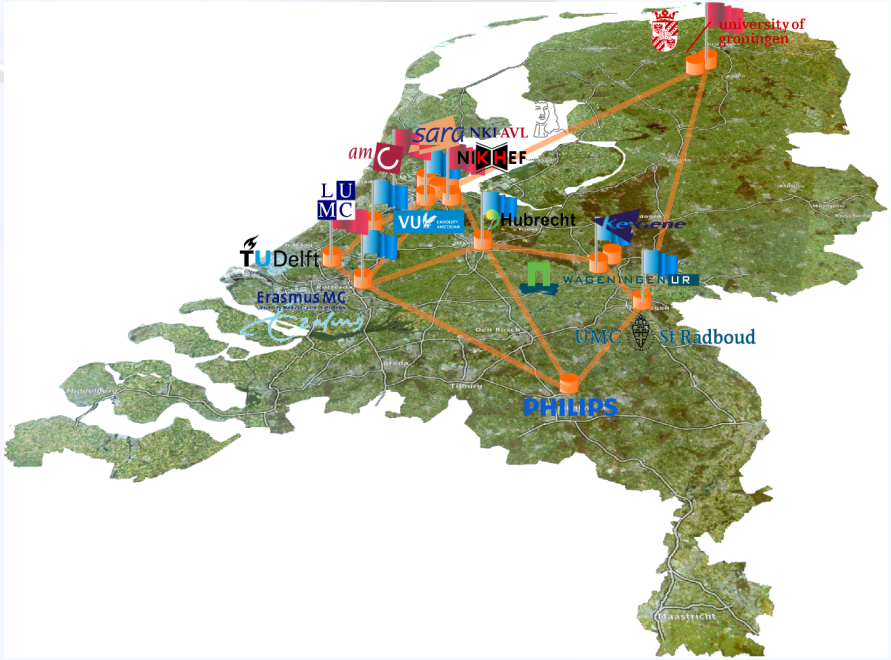
Building a central database for the research community

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

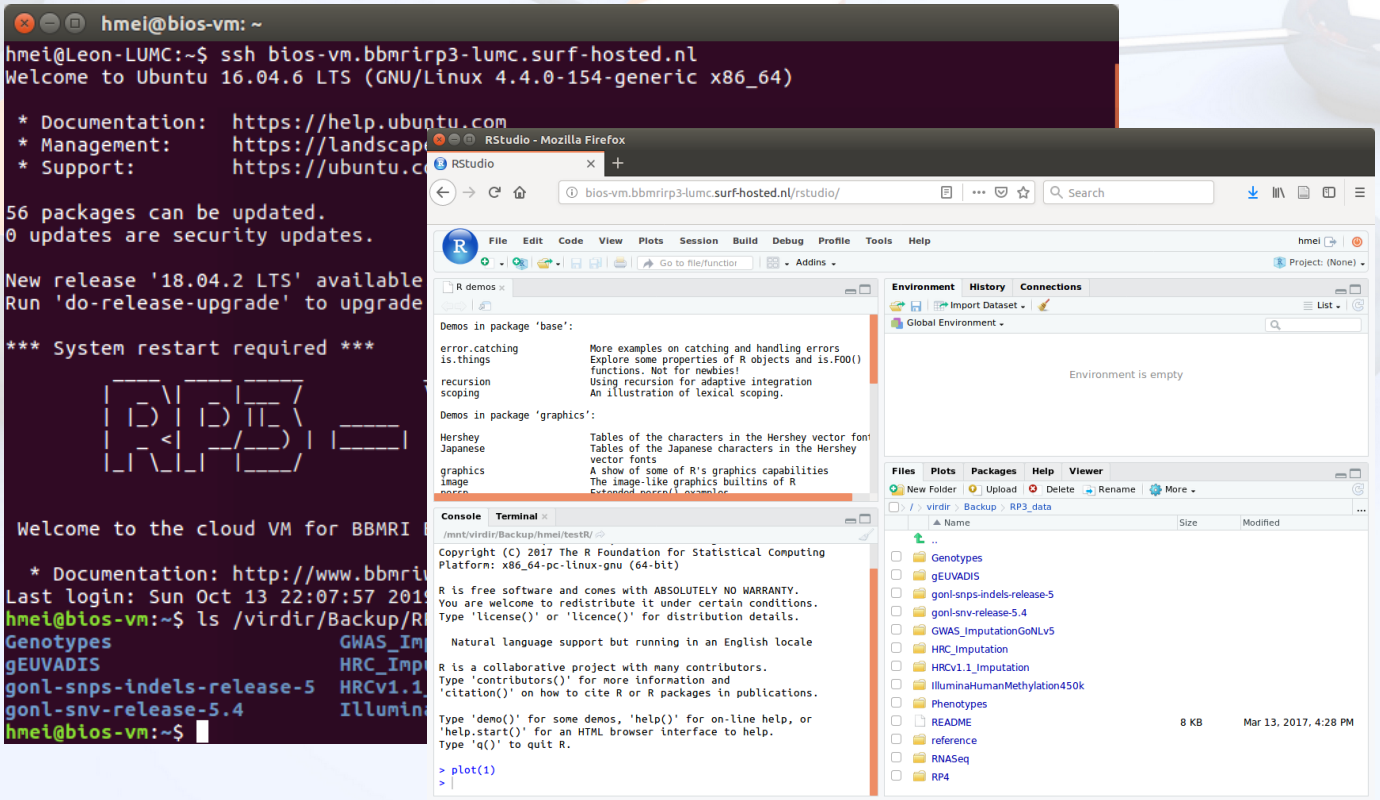
DNAseq & RNAseq pipelines







BIOS-VM



@SURFsara

HPC cloud

~50 users

64core

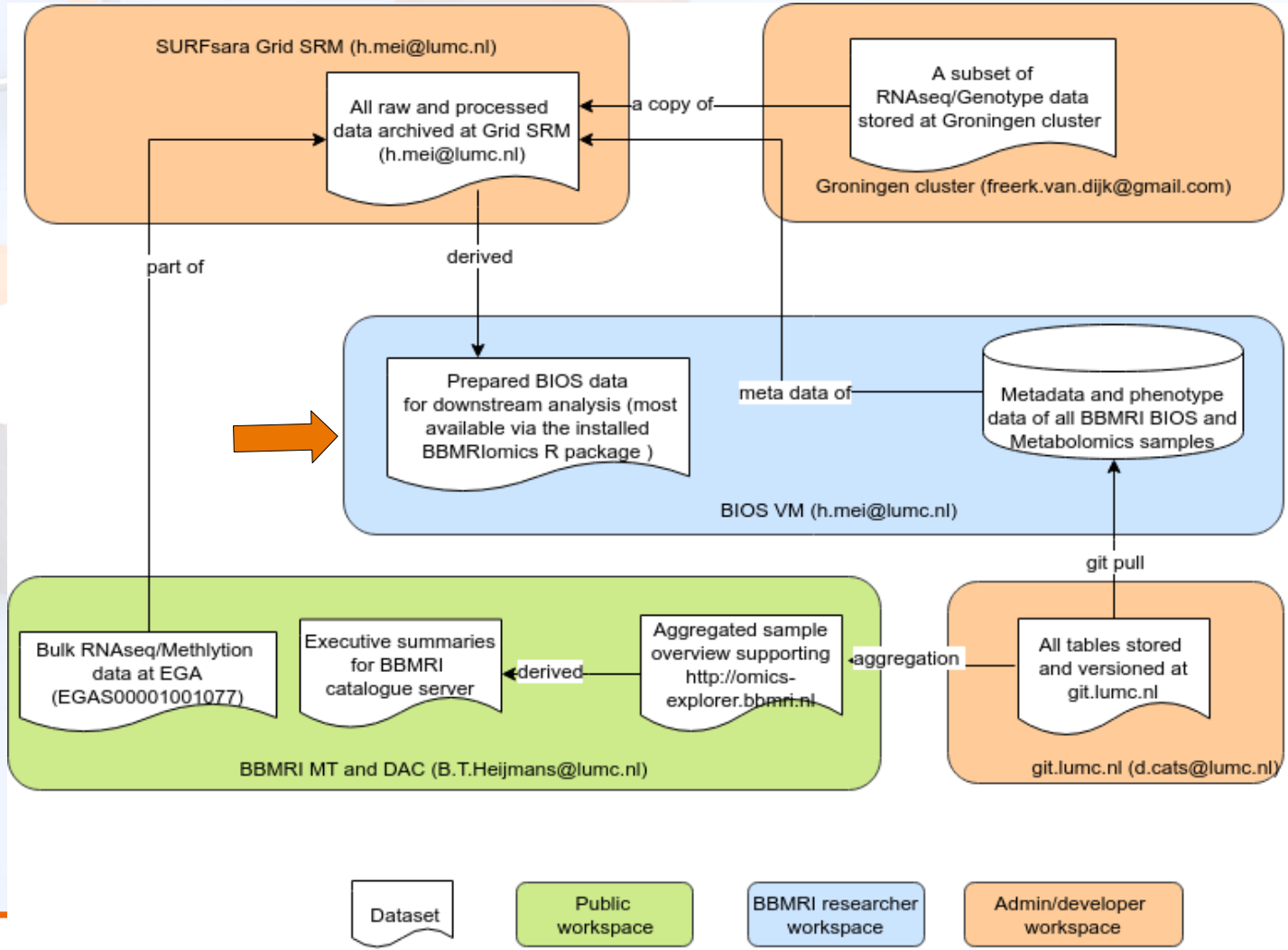
250G mem

5TB

CoC

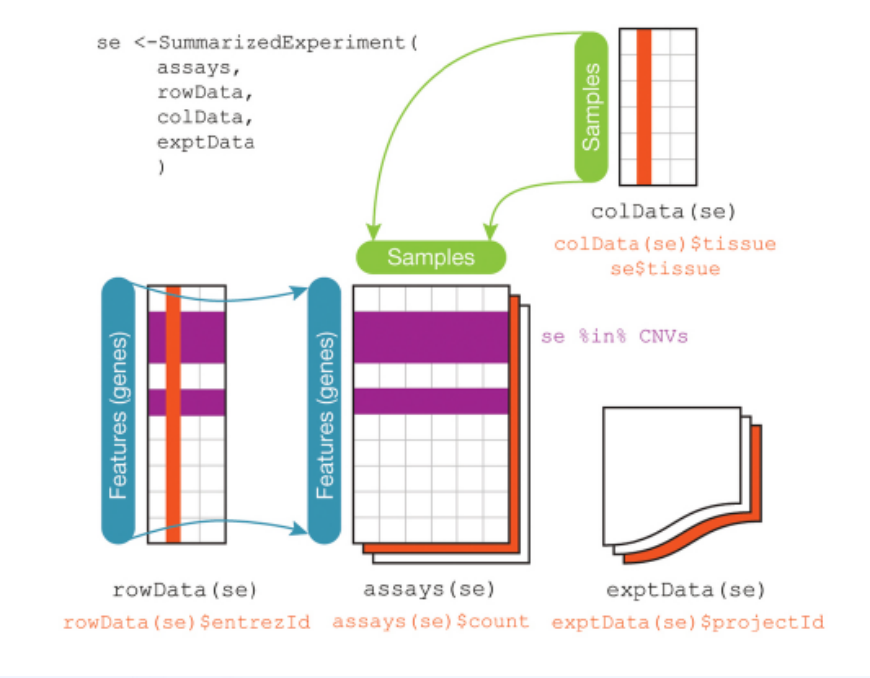
PKI auth

IP whitelisting

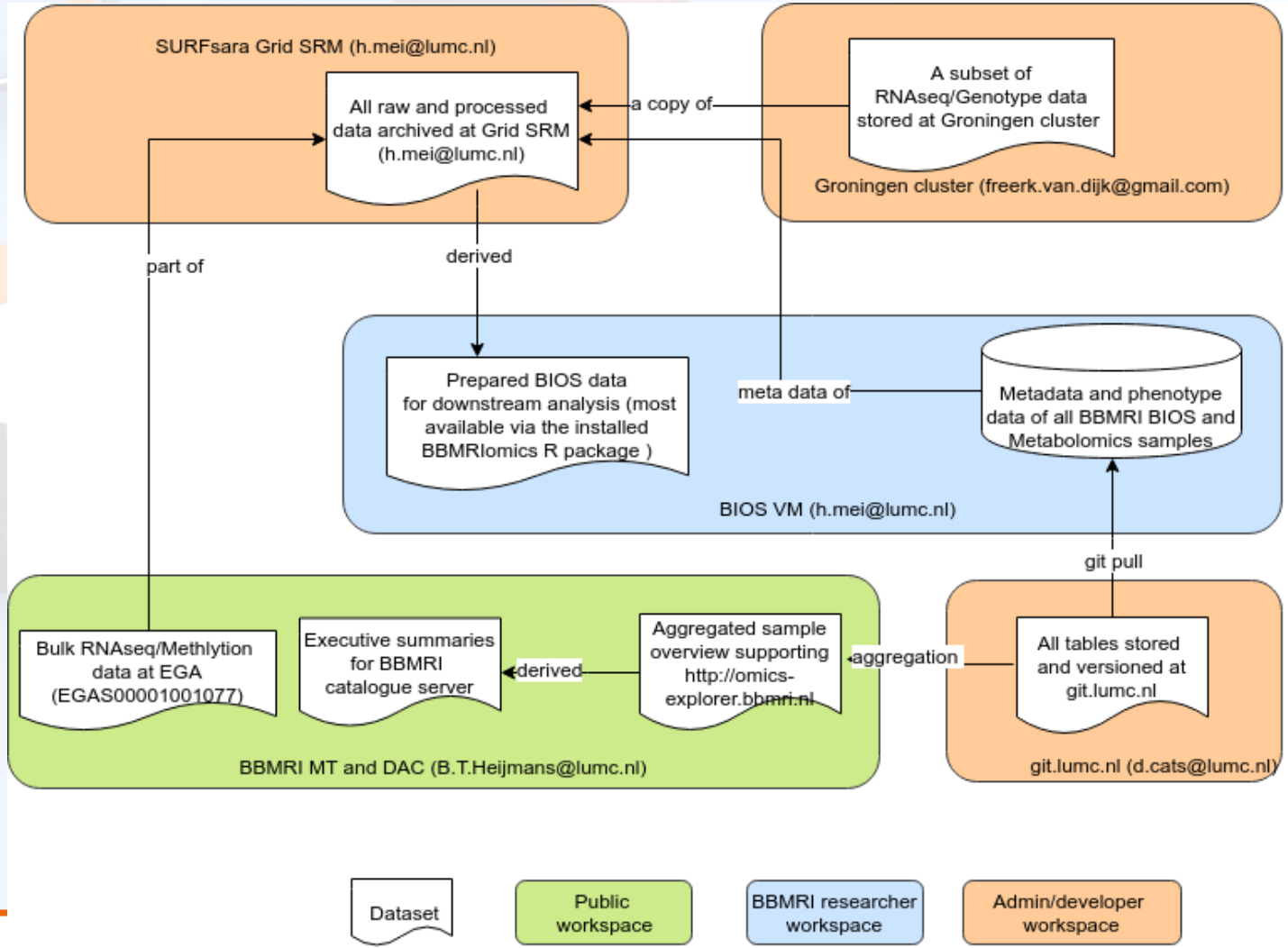

SURFsara Grid SRM (h.mei@lumc.nl)
All raw and processed data archived at Grid SRM (h.mei@lumc.nl)
a copy of
A subset of RNAseq/Genotype data stored at Groningen cluster
Groningen cluster (freerk.van.dijk@gmail.com)
part of
derived
Prepared BIOS data for downstream analysis (most available via the installed BBMRIomics R package)
meta data of
Metadata and phenotype data of all BBMRI BIOS and Metabolomics samples
BIOS VM (h.mei@lumc.nl)
git pull
Bulk RNAseq/Methlytion data at EGA (EGAS00001001077)
Executive summaries for BBMRI catalogue server
derived
Aggregated sample overview supporting http://omics-explorer.bbmri.nl
aggregation
All tables stored and versioned at git.lumc.nl
BBMRI MT and DAC (B.T.Heijmans@lumc.nl)
git.lumc.nl (d.cats@lumc.nl)
Dataset
Public workspace
BBMRI researcher workspace
Admin/developer workspace

Ready to use data

BBMRIomics: An *R* interface to BIOS data

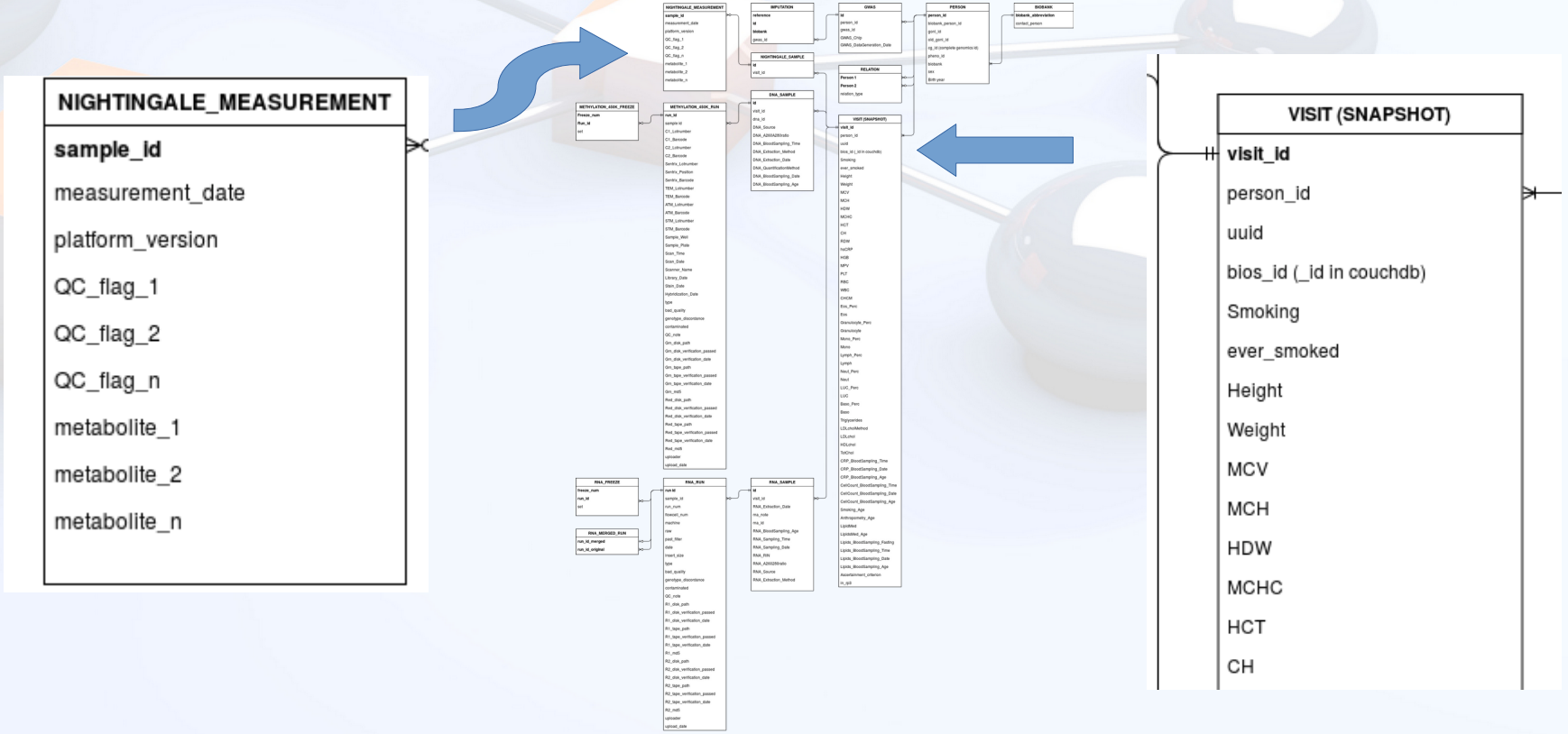


https://github.com/bbmri-nl/BBMRIomics

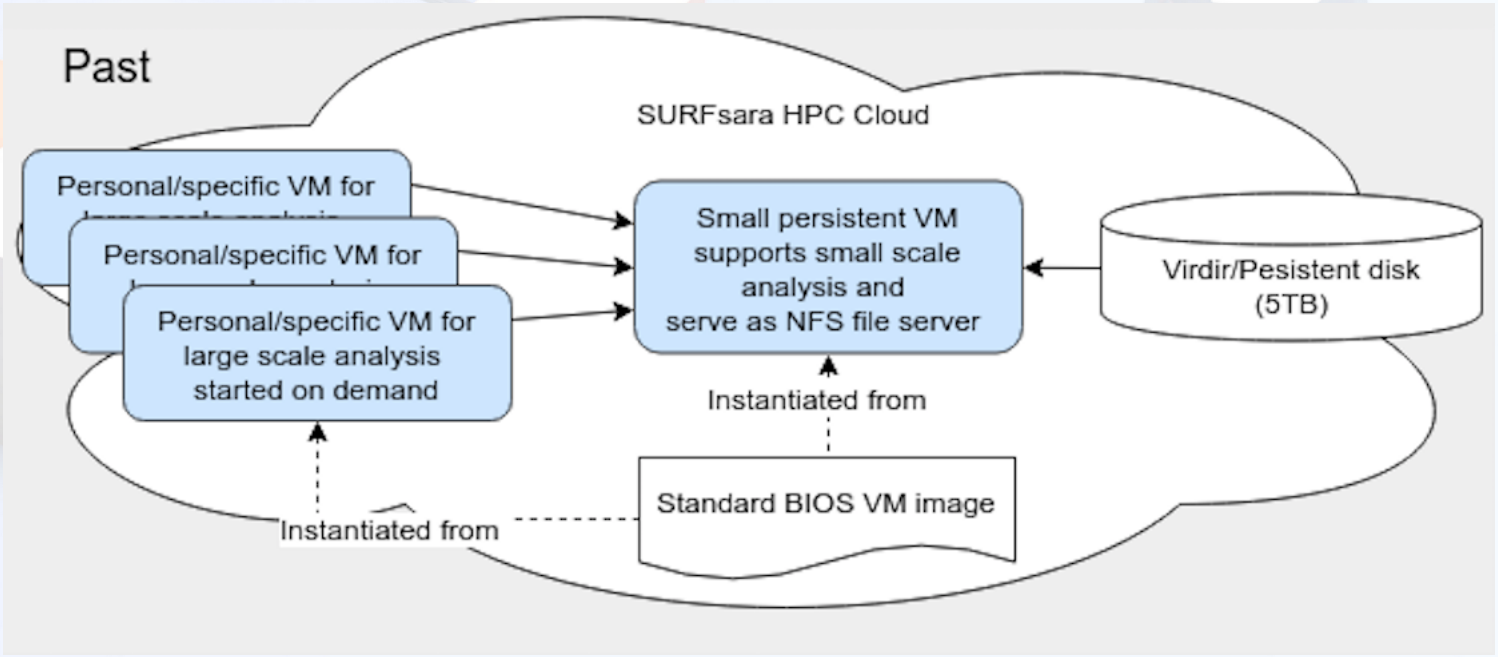

SURFsara Grid SRM (h.mei@lumc.nl)
All raw and processed data archived at Grid SRM (h.mei@lumc.nl)
a copy of
A subset of RNAseq/Genotype data stored at Groningen cluster
Groningen cluster (freerk.van.dijk@gmail.com)
part of
derived
Prepared BIOS data for downstream analysis (most available via the installed BBMRIomics R package)
meta data of
Metadata and phenotype data of all BBMRI BIOS and Metabolomics samples
BIOS VM (h.mei@lumc.nl)
git pull
Bulk RNAseq/Methlytion data at EGA (EGAS00001001077)
Executive summaries for BBMRI catalogue server
derived
Aggregated sample overview supporting http://omics-explorer.bbmri.nl
aggregation
All tables stored and versioned at git.lumc.nl
BBMRI MT and DAC (B.T.Heijmans@lumc.nl)
git.lumc.nl (d.cats@lumc.nl)
Dataset
Public workspace
BBMRI researcher workspace
Admin/developer workspace

BBMRIomics data model

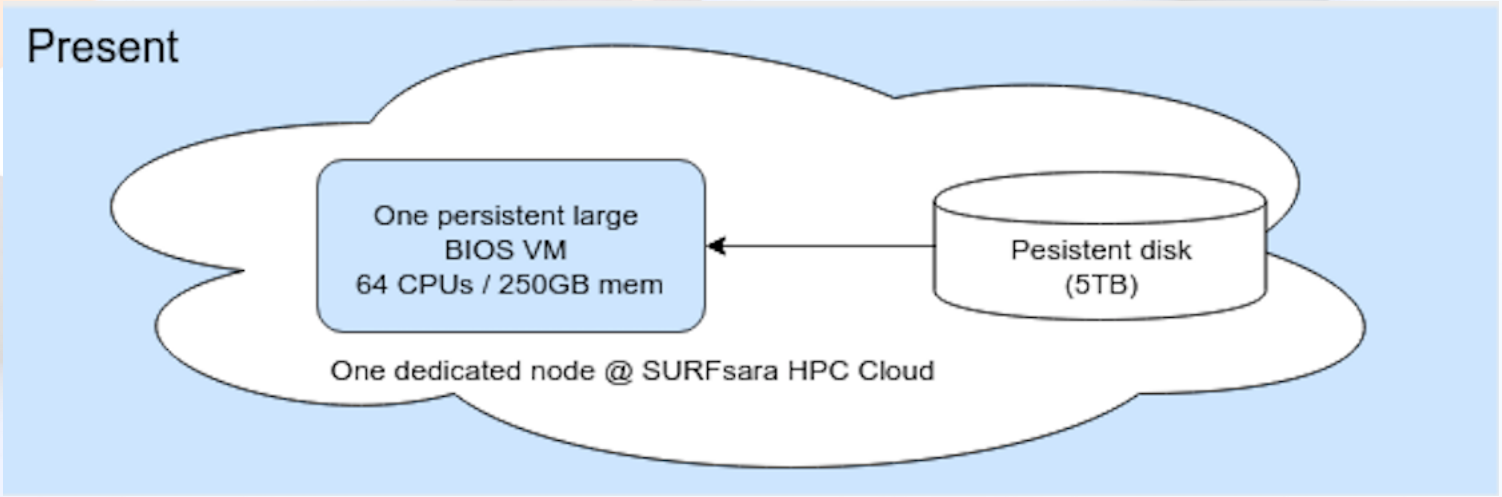
BBMRI.nl
Biobanking and BioMolecular resources Research Infrastructure The Netherlands



BIOS VM Past


Past
SURFsara HPC Cloud
Personal/specific VM for
Personal/specific VM for
Personal/specific VM for large scale analysis started on demand
Small persistent VM supports small scale analysis and serve as NFS file server
Virdir/Pesistent disk (5TB)
Instantiated from
Standard BIOS VM image
Instantiated from

BIOS VM Present

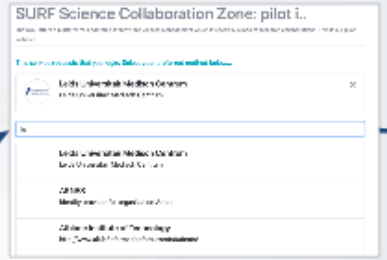


BBMRIomics workspace

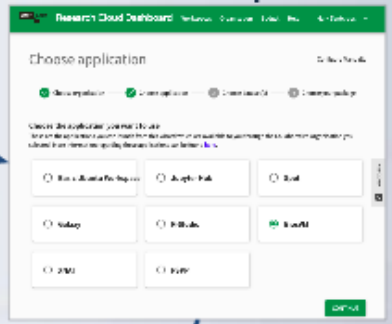
BBMRI.nl
Biobanking and BioMolecular resources Research Infrastructure The Netherlands

Homepage

Institutional Login

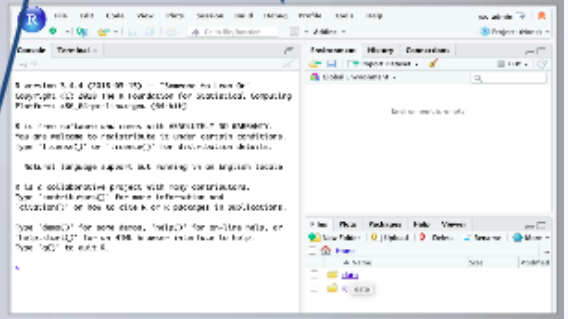


Choose workspace



Rstudio login

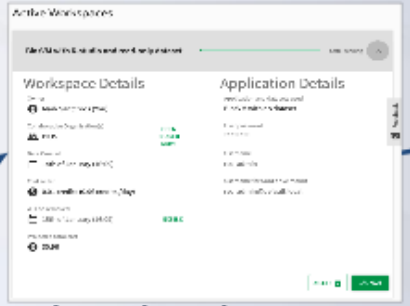
Ready for analysis!



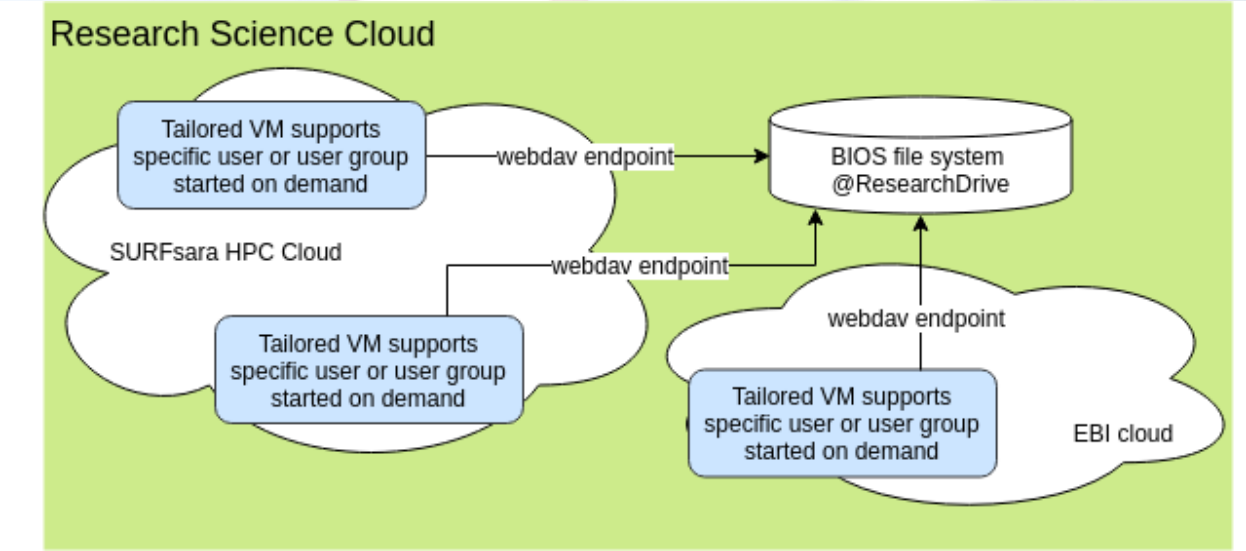
Data location

Resource dimensions

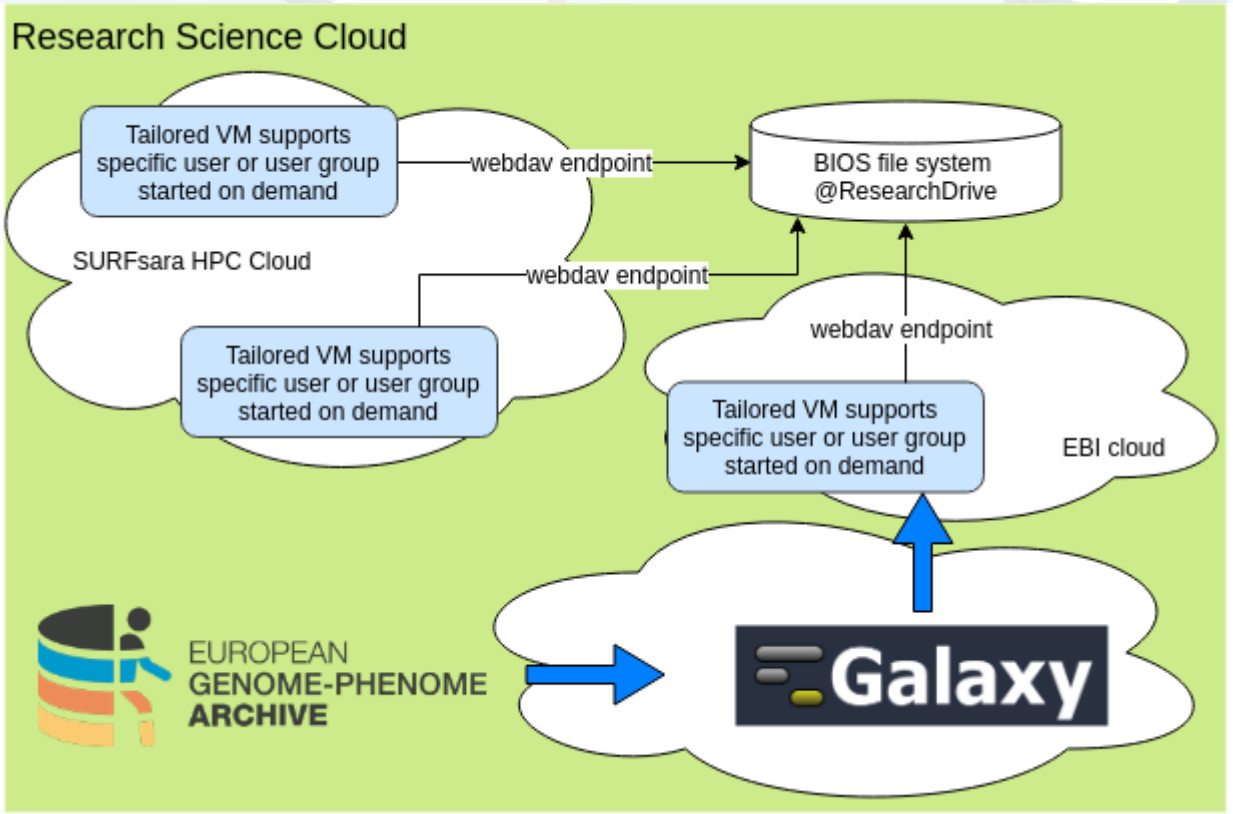
Launch workspace



BIOS VM Future



BIOS VM Future (VRE?)



BBMRI.nl
Biobanking and
BioMolecular resources
Research Infrastructure
The Netherlands

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).

BBMRI.nl
Biobanking and BioMolecular resources Research Infrastructure The Netherlands

BIOCONDA

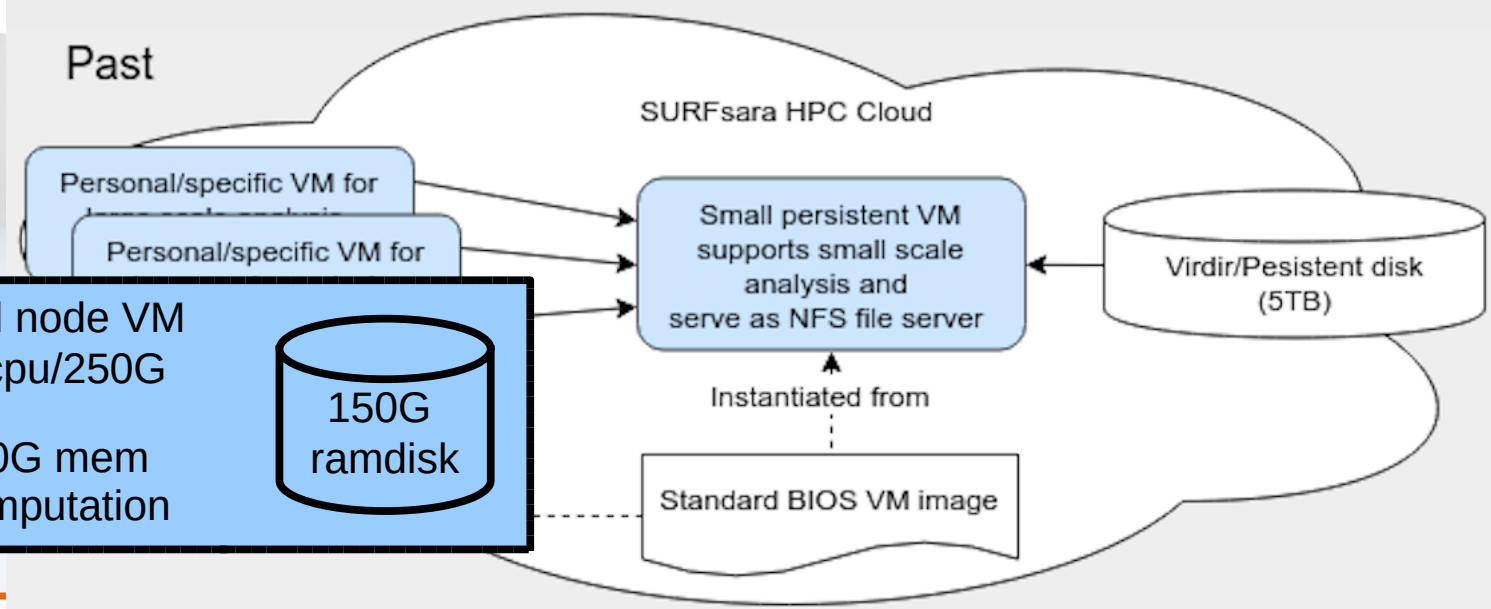
docker

Snakemake
a Python workflow manager

{wdl}

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.nl
Biobanking and BioMolecular resources Research Infrastructure The Netherlands

umcg

VU VRIJE UNIVERSITEIT AMSTERDAM

amc

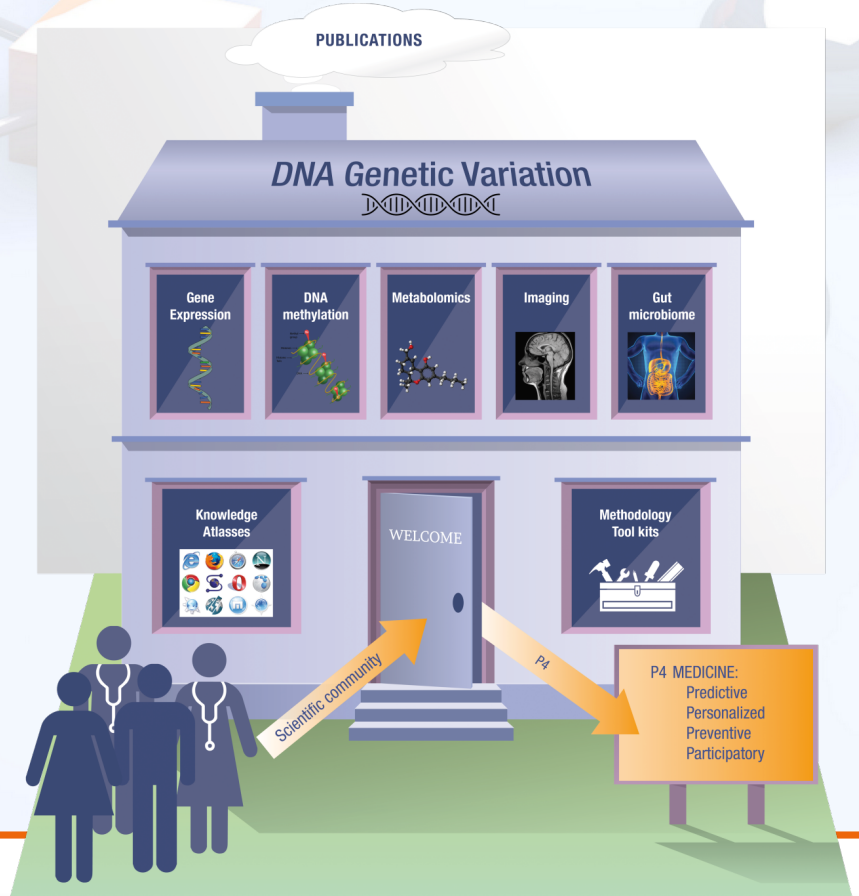
LUMC

UMC Utrecht

Erasmus MC University Medical Center Rotterdam

Radboudumc

Maastricht University



BBMRI.nl
Biobanking and BioMolecular resources Research Infrastructure The Netherlands

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.