

Importing data and tools to use in the RE

Emily Perry

Research Engagement Manager

10th February 2026



Data security

- This training session will include data from the GEL Research Environment
- As part of your IG training you have agreed to not distribute these data in any way
- You are not allowed to:
 - Invite colleagues to watch this training with you
 - Take any screenshots or videos of the training
 - Share your webinar link (we will remove anyone who is here twice)
- We are recording and will distribute the censored video later

Questions



All your
microphones
are muted



Use the Zoom
Q&A to ask
questions



Upvote your
favourite
questions: if we
are short on
time we will
prioritise those
with the most
votes

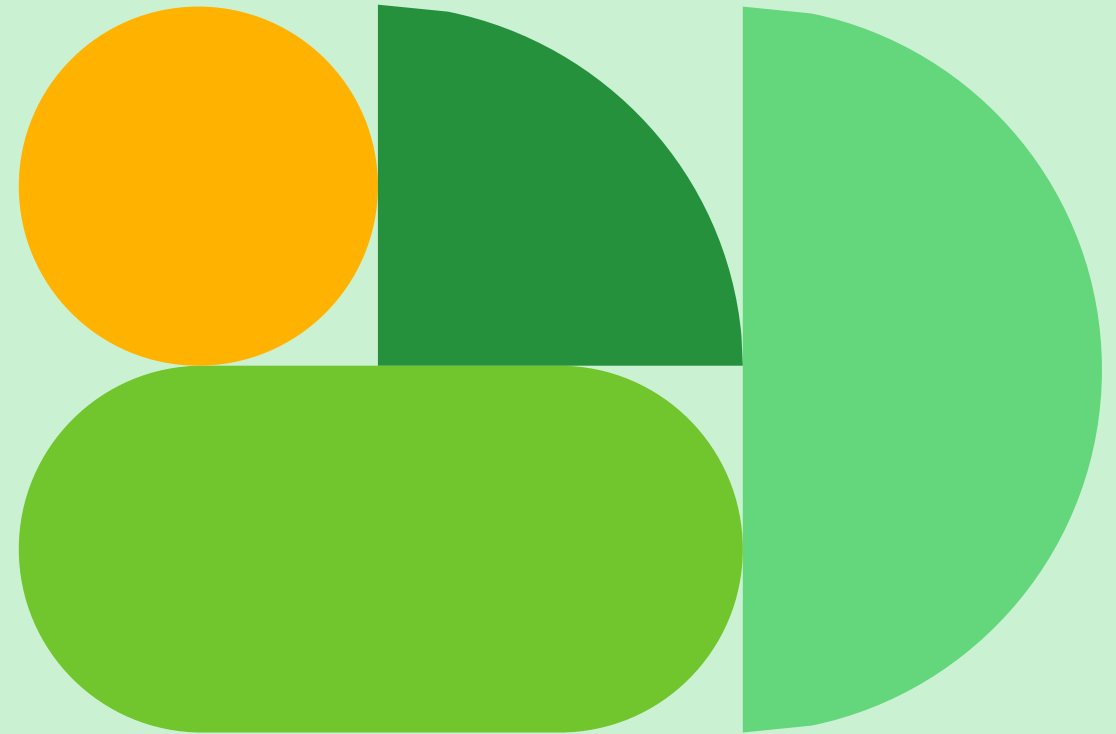
Questions



**Magdalena
Drożdż**
Bioinformatician -
Research Services

Agenda

- 1 Introduction and admin
- 2 What is already available in the RE?
- 3 Personal conda environments
- 4 Importing R packages with CRAN and Bioconductor
- 5 Importing containers with Singularity
- 6 Using Airlock to bring data/software in
- 7 Making a software request
- 8 CloudOS – importing tools and pipelines on the Cloud
- 9 Software licensing and requirements
- 10 Help and questions



2. What is already in available the RE?

Public data

1000 Genomes

[AmpliconArchitect](#)

[Annovar database](#)

Battenberg

bigzips

BLAST

CADD

Centrifuge database

Centromeres and telomeres

Clair

ClinVar

[COSMIC](#)

[dbNSFP](#)

dbSNP

deepVariant

InterPro

Ensembl

eQTLGen

ExAC

Exomiser

ExpansionHunter

FannsDB

FAVORannotator

GATK

GENCODE

GERP

Genome in a Bottle

gnomAD

GTE_x

HGDP

IGV

IMPC

Kraken

Liftover

mhcflurry

MutSigCV

NanoPlot

Nanopore WGS consortium

NCBI BLAST

Nextflow DB

GWAS Catalog

OncodriveFML

PathSeq

PhyloP 100 way

Picard

Reference genomes

R packages

Rvtests

SAIGE

SeqSeek

SGDP

Slivar

SpliceAI

TCGA PCAWG

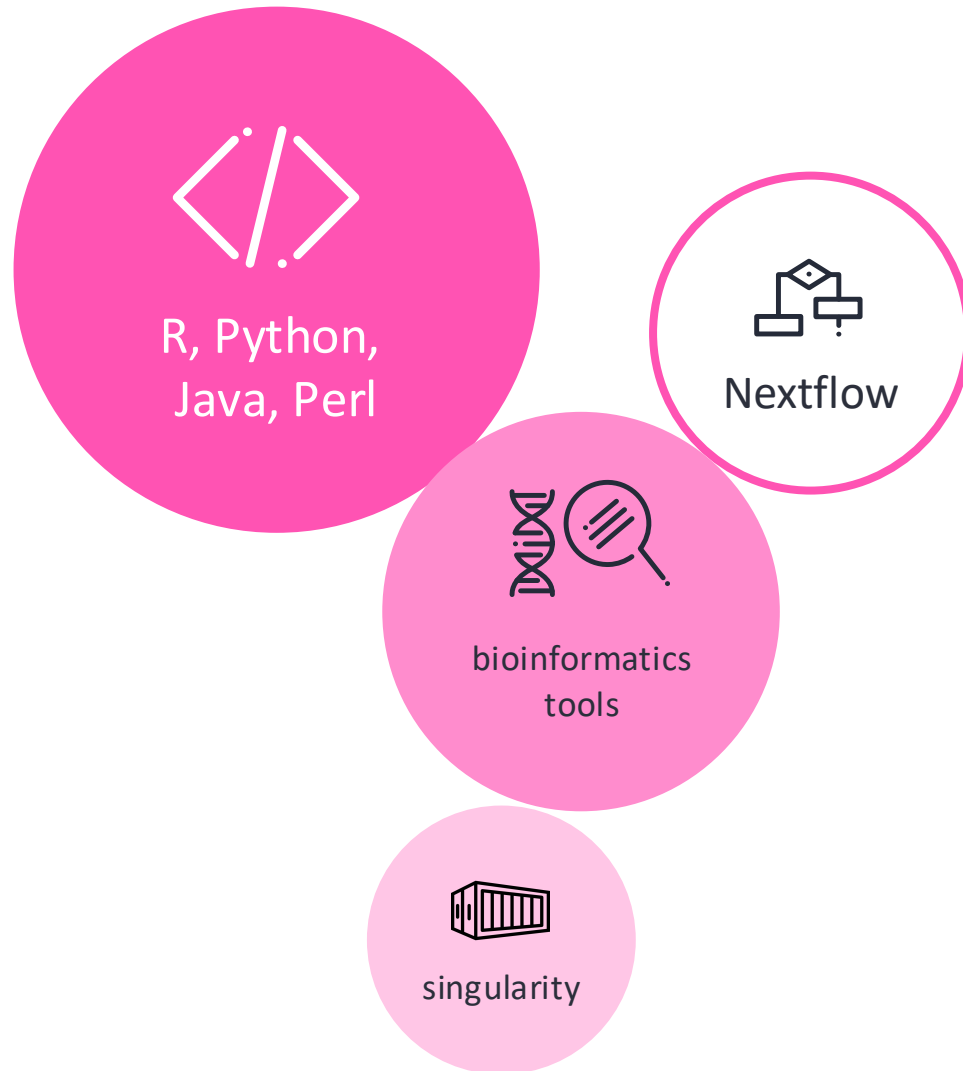
TOPMed

UCSC

VEP

wwPDB

Software on the HPC

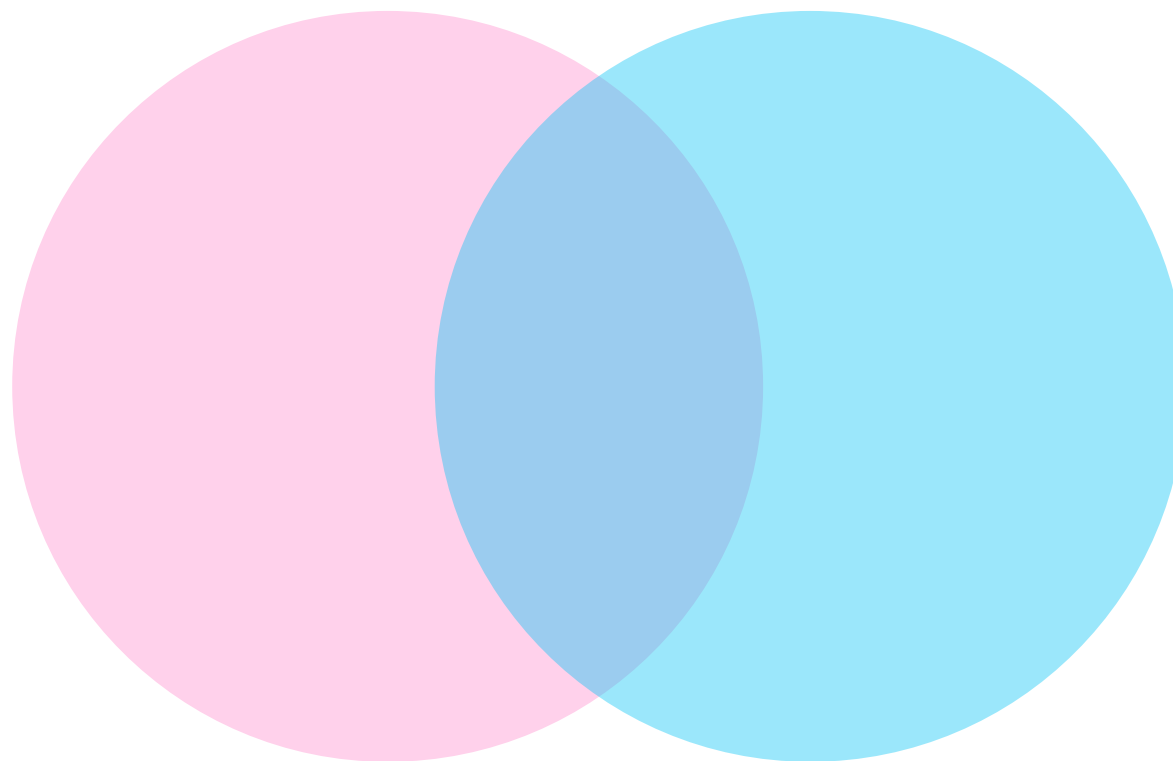


aliview/1.28
annovar/2019Nov
annovar/2024-03-14
ant/1.9.16
bamtools/2.5.2
bcftools/1.16
bcftools/1.21
beagle/5.4
bedtools/2.30.0
bedtools/2.31.1
bison/3.8.0
blat/1.0
boost/1.83
boost/1.84
cargo/0.10.5
cmake/3.24.0
cmake/ECM
conform/240524
cromwell/v65
curl/7.81.0
eigen/3.3.9
exomiser/13.3.0
exonerate/2.2.0
ExpansionHunter/5.0.0
ExpansionHunterDenovo/0.9.0
fetc/1.9.3
flex/2.6.4
gatk/4.5.0.0
gcc/10.4.0
gdal/3.7.0
geos/3.12.2
gmp/6.2.1
go/1.25.0
gsl/2.7
haplocheck/1.3.3
hipstr/0.7
hla-la/1.0.3

hmftools/2024-02-06
htslib/1.18
htslib/1.21
inkscape/1.3
java/1.8
java/11.0.2
java/17.0.2
java/19.0.2
java/21.0.2
java/23.0.1
jq/1.7.1
king/2.3.2
kraken/1.1.1
kraken2/2.1.3
libdeflate/1.20
libffi/3.4.7
libgit2/1.6.2
libpcre/8.45
libpng/1.6.43
libtiff/4.7.0
locuszoom/1.4
m4/1.4.19
magma/1.10
maven/3.9.6
miniconda/23.11.0
mpc/1.3.1
MPFR/4.2.0
ncurses/6.4
nextflow/22.10.5
nextflow/23.04
nextflow/23.10
nextflow/23.10-with-plugins
nextflow/24.04.2-with-plugins
nf-core/0.3.1
nf-test/0.7.3
nf-test/0.8.2
nf-test/0.9.0

openrefine/3.7.4
openssl/1.1.1o
openssl/1.1.1t
openssl/3.6.0
perl/5.38.2
picard/3.1.1
plink/1.9
plink/2.0
plink/2.00a3.3LM
postgresql/17.4
proj/8.2.1
python/3.11
python/3.13.9
qgrs-cpp/1.0
qt/6.10.1
R/4.3.3
R/4.5.1
regenie/3.4.1
REViewer/0.2.7
rtg-tools/3.12.1
rust/1.91.1
samtools/1.16.1
samtools/1.21
singularity/3.8.3
singularity/4.1.1
singularity/4.1.1-test3
sniffles/1.0.11
sqlite3/3.38.1
strelka/2.9.10
tabix/1.21
udunits/2.2.28
xz/5.4.7
yt-dlp/2025-12-08
zlib/1.3

R packages

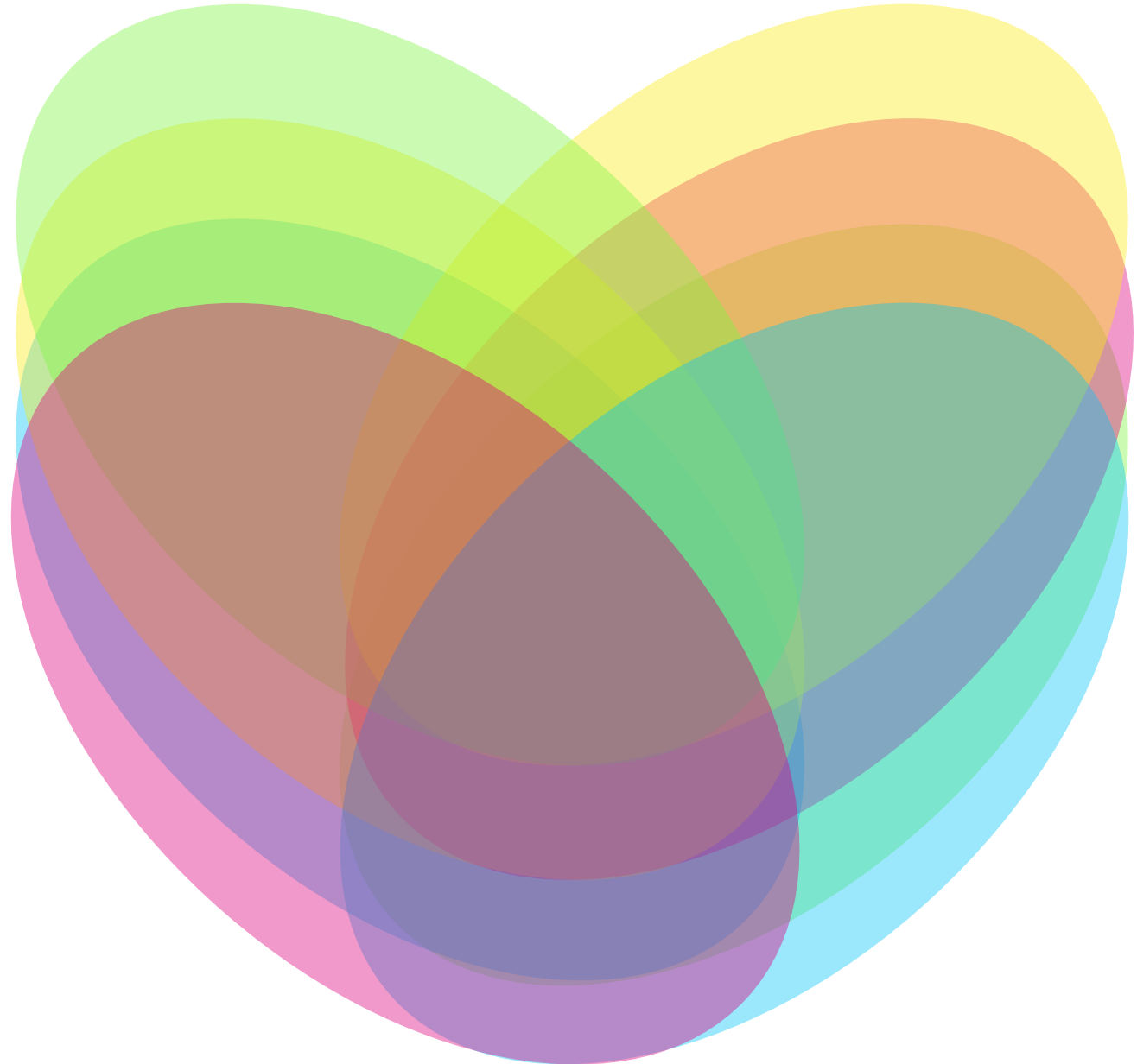


R/4.5.1

R/4.3.3

Python packages in conda environments

idppy3
idppy2
ipy3nopyrev1
ipy3pypirev1
ipy3nopypi
ipy3pypi
ipy3tf2
ipy3keras
ipy3mavis
py3nopypirev1
py3nopypi
py3pypi
py2_7_12nopypi
py2_7_12pypi
idpcorepy3_6_5rev1
idpcorepy2_7_12rev1
idpcorepy3tf2rev1



Existing data/tools demo



-  Home
-  Panel App
-  Visual Studio Code
-  Airlock
-  Participant Explorer
-  Welcome Pack
-  Ensembl
-  R
-  IGV Browser
-  RE Docs
-  IVA
-  Research Registry
-  Labkey
-  RStudio
-  Open Targets
-  Stata

3. Personal conda environments

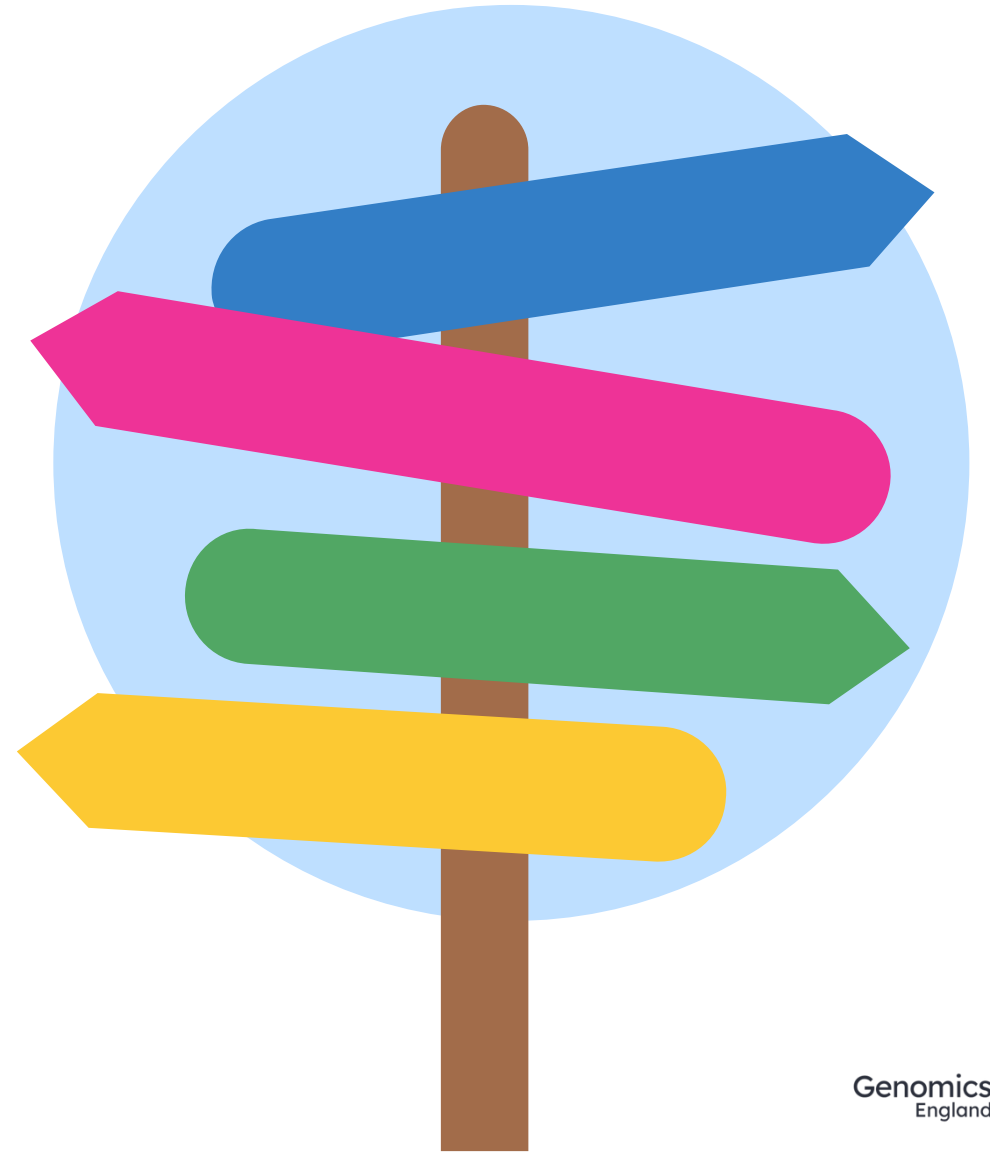


Authorised channels

anaconda

bioconda

conda-forge



Process

Copy config file to your folder:

```
cp /gel_data_resources/example_config_files/Helix/.condarc ~/.
```

Create your environment

```
conda create python==<version_number> --prefix  
/path/to/env/location
```

Use the env path to activate:

```
source /resources/conda/miniconda3/bin/activate conda activate  
/path/to/env/location
```

Use pip to install further packages using a proxy path:

```
pip install <package_name> --index-url  
https://artifactory.aws.gel.ac/artifactory/api/pypi/pypi/simple
```

Conda demo



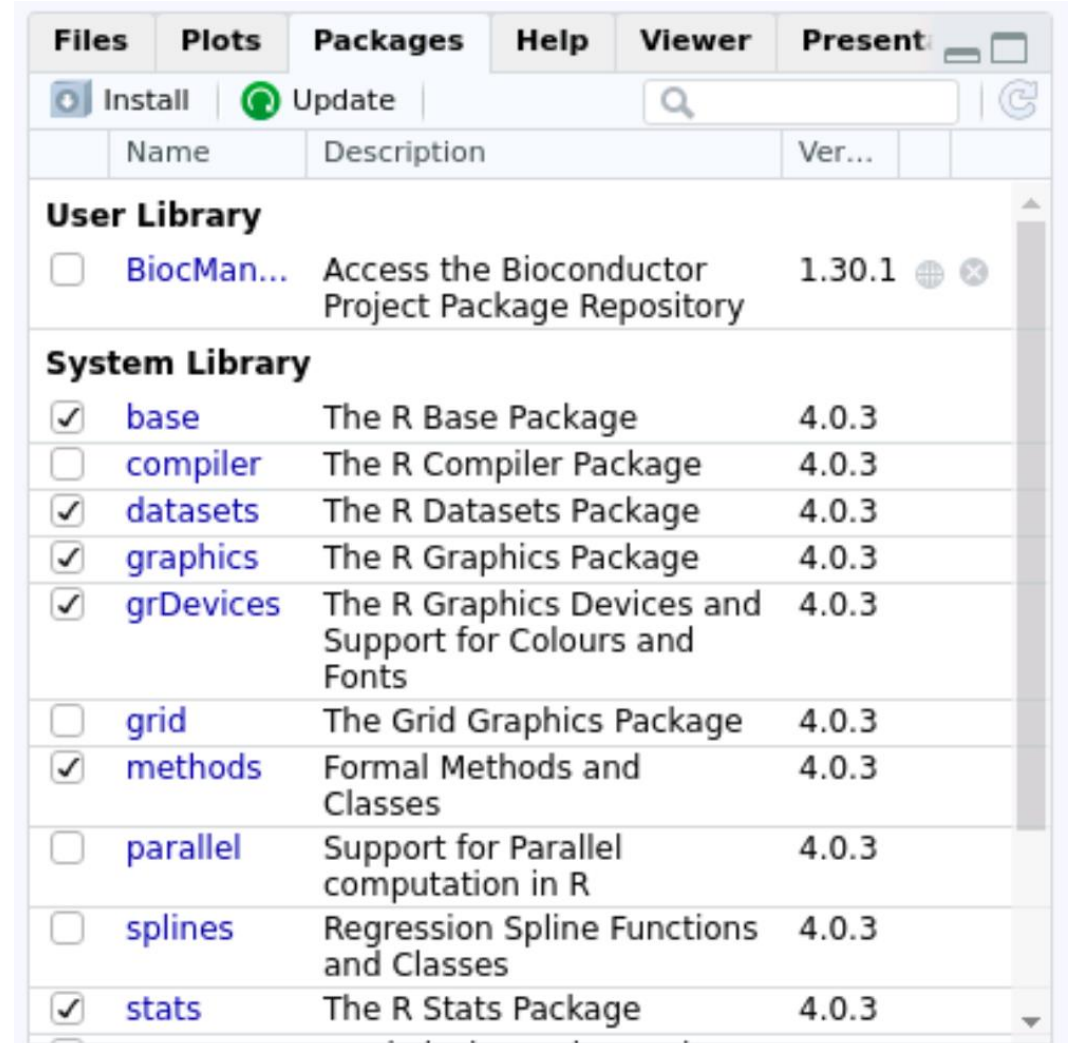
- Home
- Panel App
- Visual Studio Code
- Airlock
- Participant Explorer
- Welcome Pack
- Ensembl
- R
- IGV Browser
- RE Docs
- IVA
- Research Registry
- Labkey
- RStudio
- Open Targets
- Stata

```
eperry@U-3U0LARZSM2ZGF: ~  
eperry@hpc-doublehelix-login-053359cc:~$
```

4. Importing R packages with CRAN and Bioconductor

Loading libraries

```
library(library_name)
```



The screenshot shows the RStudio interface with the 'Packages' tab selected. It displays a list of installed and available packages, categorized into 'User Library' and 'System Library'. The 'BiocManager' package is listed under the 'User Library' and is not installed. The 'System Library' contains several base R packages, most of which are installed (checked).

	Name	Description	Ver...
User Library			
☐	BiocMan...	Access the Bioconductor Project Package Repository	1.30.1
System Library			
☒	base	The R Base Package	4.0.3
☐	compiler	The R Compiler Package	4.0.3
☒	datasets	The R Datasets Package	4.0.3
☒	graphics	The R Graphics Package	4.0.3
☒	grDevices	The R Graphics Devices and Support for Colours and Fonts	4.0.3
☐	grid	The Grid Graphics Package	4.0.3
☒	methods	Formal Methods and Classes	4.0.3
☐	parallel	Support for Parallel computation in R	4.0.3
☐	splines	Regression Spline Functions and Classes	4.0.3
☒	stats	The R Stats Package	4.0.3

Bioconductor

.Rprofile

```
options(  
  BIOCONDUCTOR_CONFIG_FILE =  
  "https://artifactory.aws.gel.ac:443/artifactory/bioconductor.org-  
    cache/config.yaml"  
)
```

console

```
.libPaths(c( .libPaths(), "/tools/aws-workspace-apps/ce/R/4.3.3"))  
library("BiocManager")  
BiocManager::install("<package_name>")  
library(<package_name>)
```

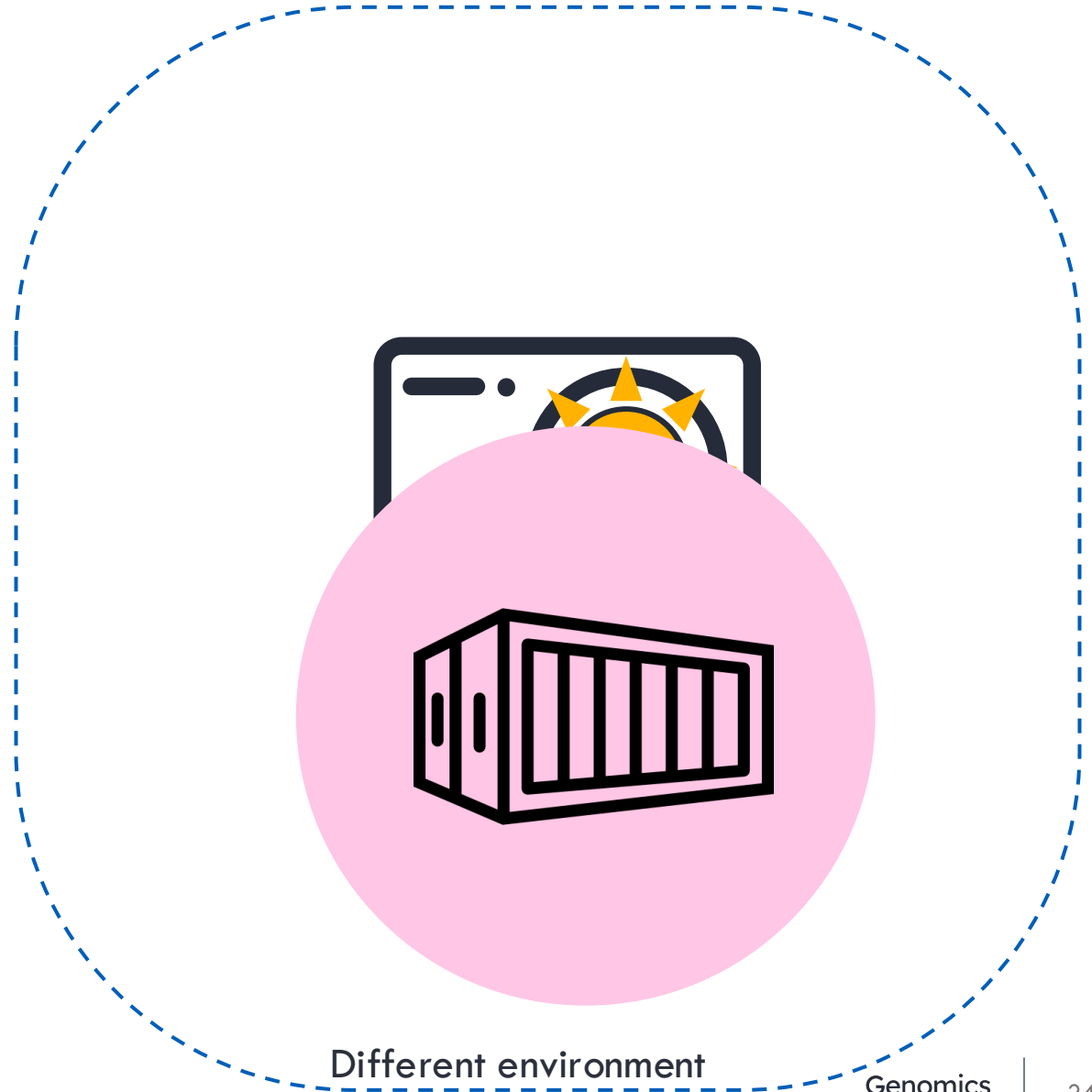
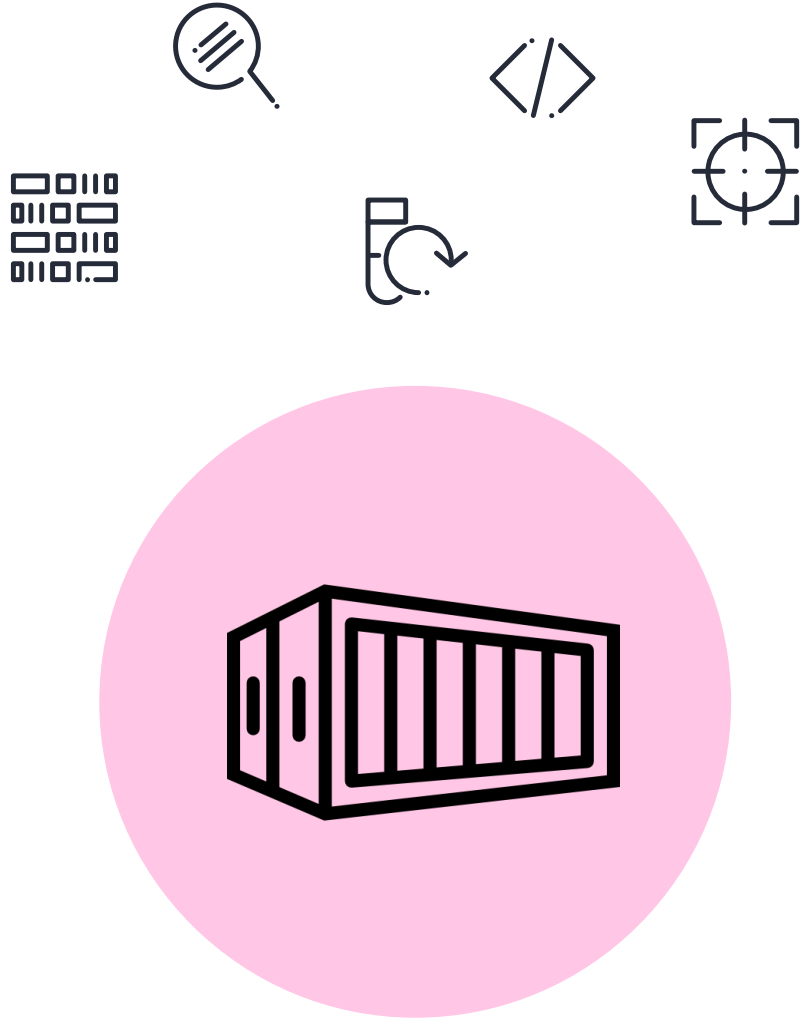
CRAN/Bioconductor demo



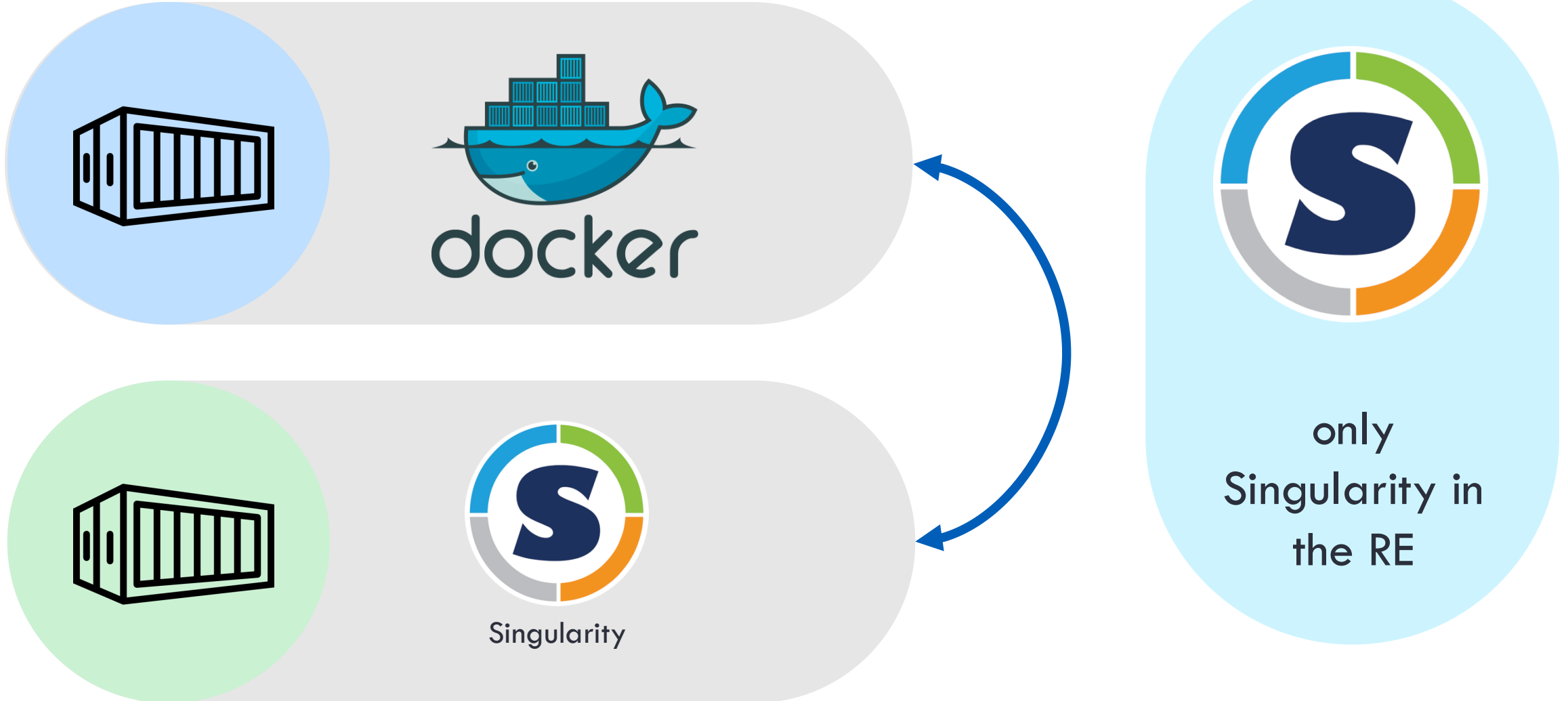
- Home
- Open Targets
- RStudio
- Airlock
- Panel App
- Stata
- Ensembl
- Participant Explorer
- Visual Studio Code
- IGV Browser
- R
- Welcome Pack
- IVA
- RE Docs
- Labkey
- Research Registry

5. Importing containers using Singularity

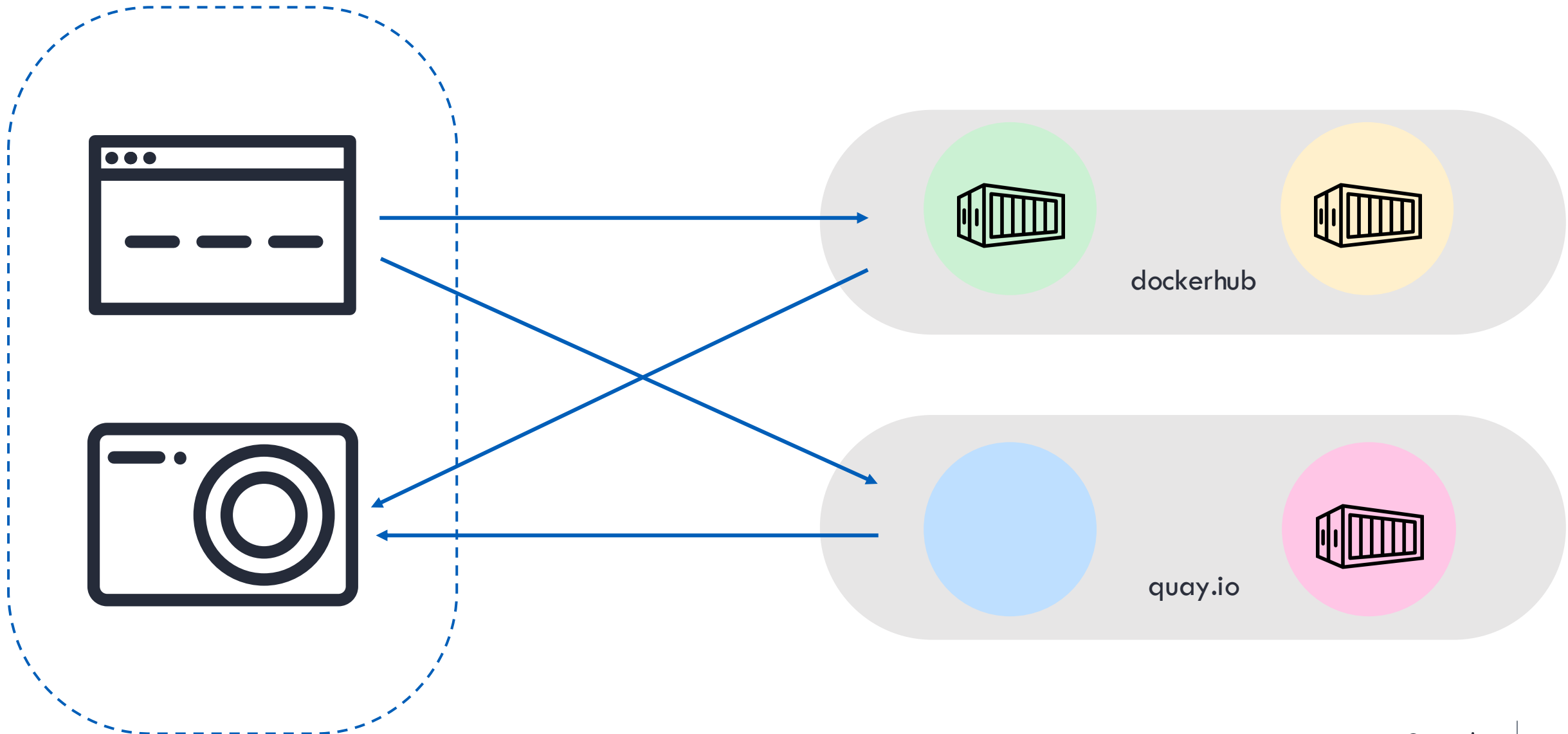
Containers



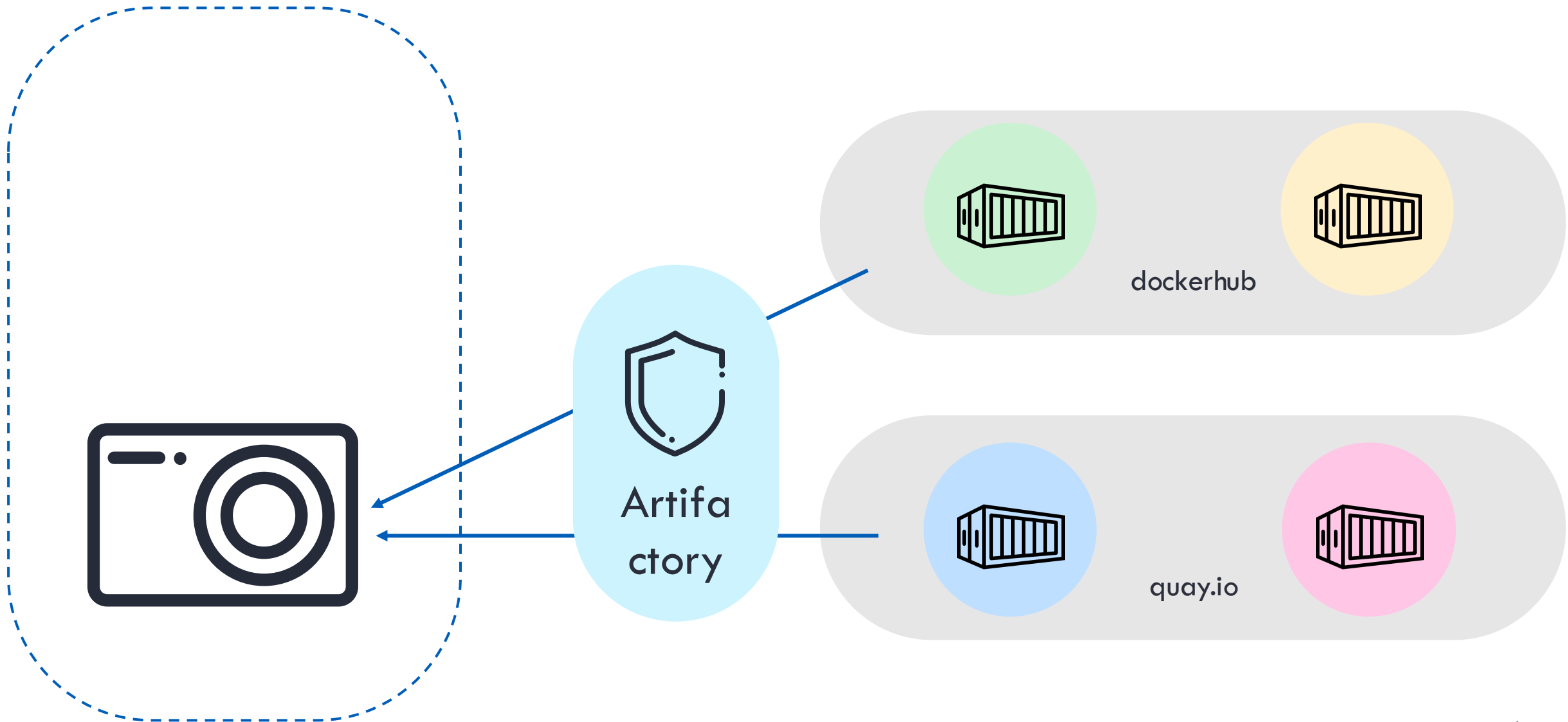
Container types



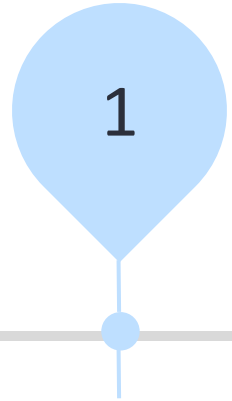
Container repositories



Containers in the RE



Steps to import container



Find container in repository: identify docker or quay.io container location



modify import location to include artifactory redirect



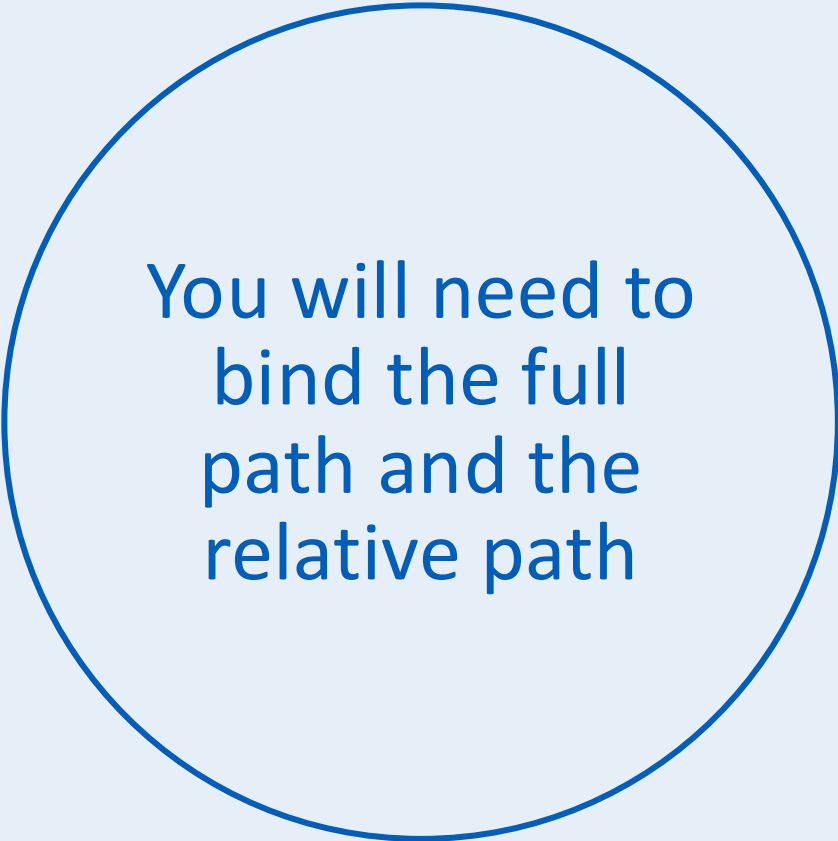
pull container from an interactive job on the HPC



mount data and run analysis

Mounting data

```
--bind  
/nas/weka.gel.zone/pgen_genomes:/nas/weka.gel.zone/p  
gen_genomes --bind /genomes:/genomes  
--bind  
/nas/weka.gel.zone/pgen_int_data_resources:/nas/weka  
.gel.zone/pgen_int_data_resources --bind  
/gel_data_resources:/gel_data_resources  
--bind  
/nas/weka.gel.zone/pgen_public_data_resources:/nas/w  
eka.gel.zone/pgen_public_data_resources --bind  
/public_data_resources:/public_data_resources  
/nas/weka.gel.zone/re_scratch:/nas/weka.gel.zone/re_  
scratch --bind /re_scratch:/re_scratch  
--bind  
/nas/weka.gel.zone/re_gecip:/nas/weka.gel.zone/re_ge  
cip --bind /re_gecip:/re_gecip  
--bind  
/nas/weka.gel.zone/discovery_forum:/nas/weka.gel.zon  
e/discovery_forum --bind  
/discovery_forum:/discovery_forum
```



You will need to
bind the full
path and the
relative path

Singularity demo



```
eperry@U-3U0LARZSM2ZGF: ~  
eperry@U-3U0LARZSM2ZGF:~$ ssh eperry@doublehelix.helix.preprod.aws.gel.ac  
(eperry@doublehelix.helix.preprod.aws.gel.ac) Password:  
  
*****  
*****  
**                               Welcome to the Genomics England HPC (Double  
e Helix) preprod Environment  
**  
**  
**  
** This environment is mainly for Genomics England Quality & Assurance team to c  
heck and run test suites  
**  
**  
** Thank you!  
**  
*****  
*****  
  
Last login: Tue Jan 20 08:05:19 2026 from 172.21.128.217  
eperry@hpc-doublehelix-login-06f658c0:~$
```



Home



Panel App

Visual Studio
Code

Airlock

Participant
Explorer

Welcome Pack



Ensembl



R



IGV Browser



RE Docs



IVA



Research Registry



Labkey



RStudio



Open Targets



Stata



6. Using Airlock to bring data/tools in

The Airlock

Data in the RE

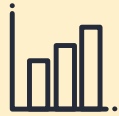


Outside world

What you can import



Scripts and tools for analysing GEL data



Data you want to compare to GEL data



Data you have consent to use



Tools you are licensed to use

Airlock demo

Camtasia 2024FileEditModifyTextViewExportWindowHelp

Homepage - Genomics Engla x

re-docs.genomicsengland.co.uk

Raise A TicketRE docsJIRA ConfluenceMulti Factor AuthChange PasswordbobSharePointbrs_gitGEL AcademyregistryBRS_meetingsbrandingstyle guidelifebit_docsmatrixslidesNotion

Genomics England Research Environment User Guide

Search

HomeGetting startedHow-to guidesDataDesktop applicationsCloudOSHigh Performance Cluster (HPC)Workflows, scripts and containersGetting helpData security and ExportTraining

Home

Genomics England Research Environment

→ Getting started

How-to guides

Data in the Research Environment

Desktop applications in the Research Environment

High Performance Cluster (HPC)

Workflows, scripts and containers

Data security and Airlock

Training

December 12, 2024

ANNOUNCEMENTS

Current data release:

- 100kGP: [/main-programme/main-programme_v19_2024-10-31](#)
- NHS GMS: [nhs-gms/nhs-gms-release_v4_2024-08-22](#)

20th December - Secondary clinical data for NHS GMS

We now have [medical record data](#) for [NHS GMS](#) participants.

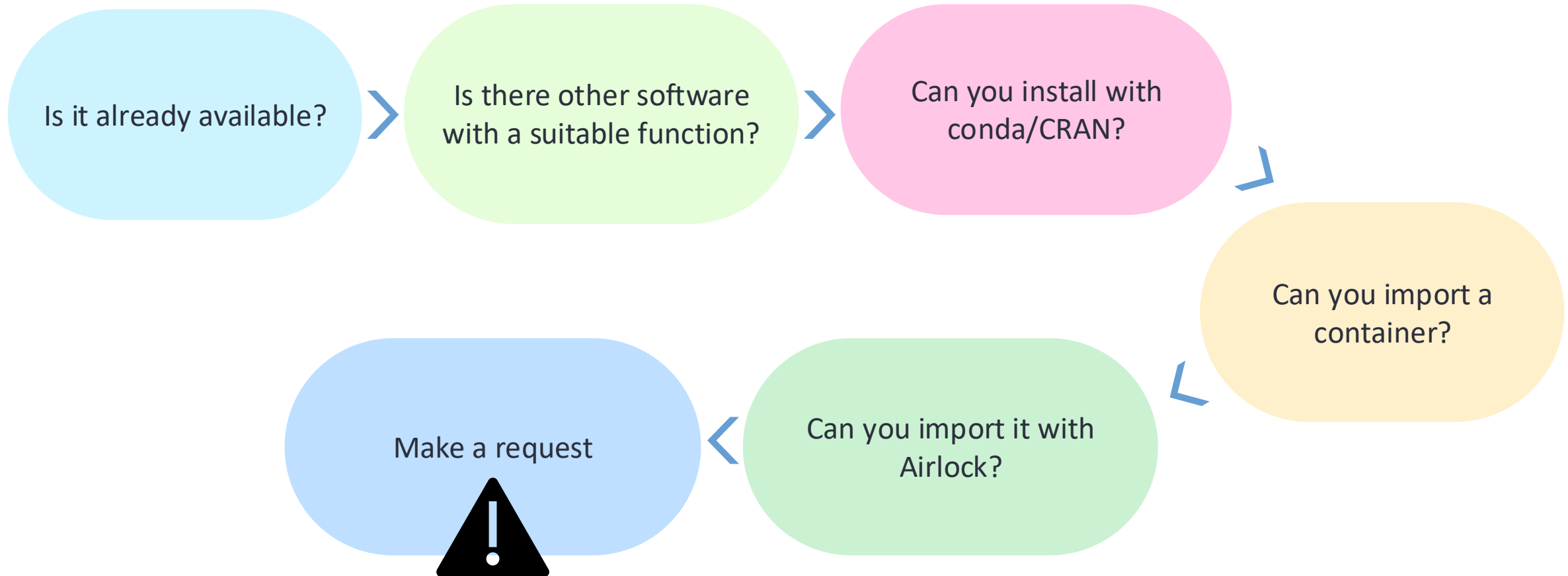
18th November - change to HPC address

The HPC address has changed to [doublehelix.helix.prod.aws.gel.ac](#).

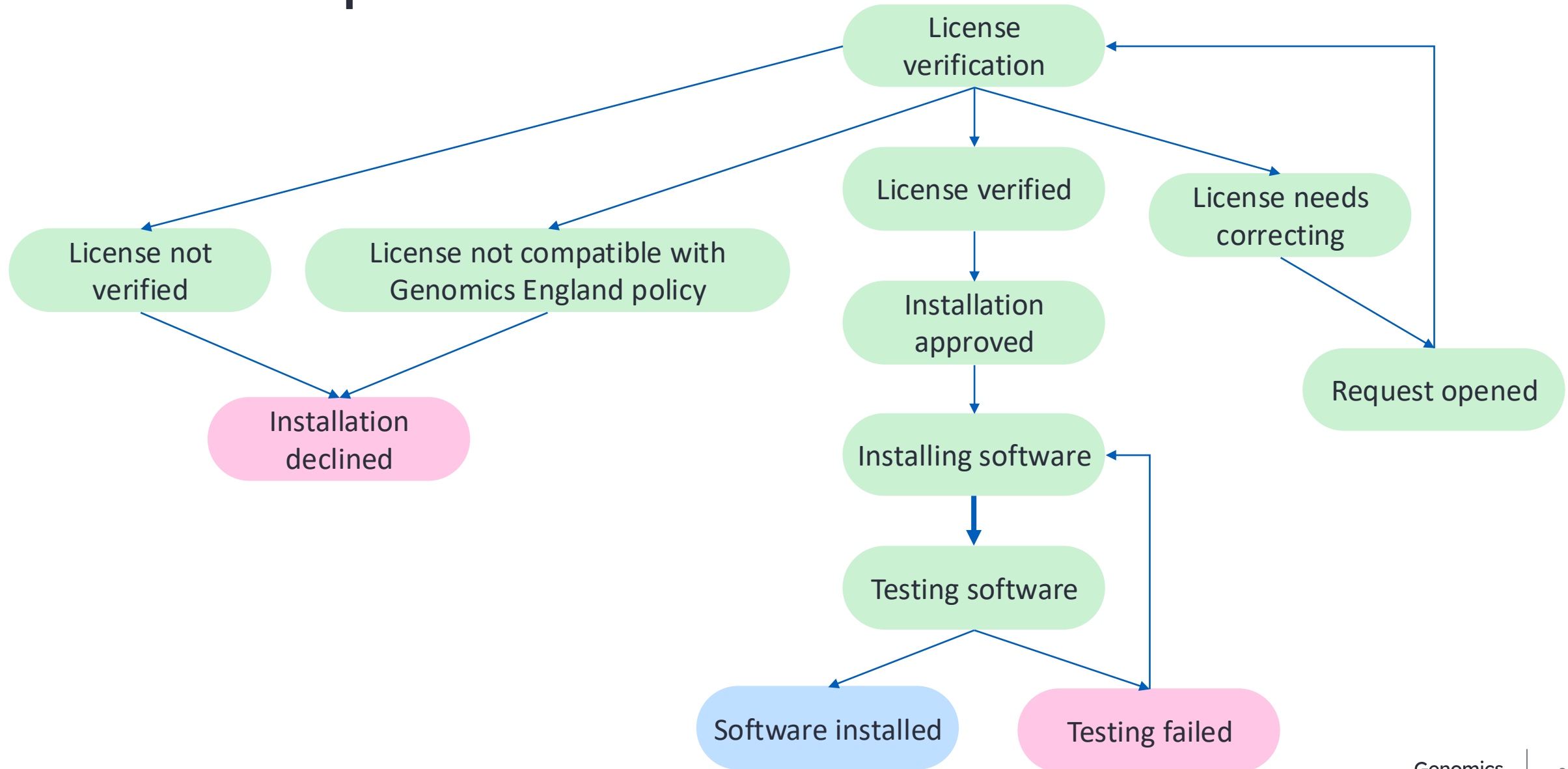
6 November 2024 - Workflow update

7. Make a software request

When to make a request



Installation process



Request demo

Camtasia 2024FileEditModifyTextViewExportWindowHelp

Homepage - Genomics Engla x

re-docs.genomicsengland.co.uk

Raise A TicketRE docsJIRA ConfluenceMulti Factor AuthChange PasswordbobSharePointbrs_gitGEL AcademyregistryBRS_meetingsbrandingstyle guidelifebit_docsmatrixslidesNotion

Genomics England Research Environment User Guide

Search

HomeGetting startedHow-to guidesDataDesktop applicationsCloudOSHigh Performance Cluster (HPC)Workflows, scripts and containersGetting helpData security and ExportTraining

Home

Genomics England Research Environment

→ Getting started

📖 How-to guides

📄 Data in the Research Environment

✂ Desktop applications in the Research Environment

🚀 High Performance Cluster (HPC)

🔗 Workflows, scripts and containers

🛡 Data security and Airlock

🎓 Training

🕒 December 12, 2024

ANNOUNCEMENTS

Current data release:

• 100kGP: [/main-programme/main-programme_v19_2024-10-31](#)

• NHS GMS: [nhs-gms/nhs-gms-release_v4_2024-08-22](#)

20th December - Secondary clinical data for NHS GMS

We now have [medical record data](#) for [NHS GMS](#) participants.

18th November - change to HPC address

The HPC address has changed to [doublehelix.helix.prod.aws.gel.ac](#).

6 November 2024 - Workflow update

8. CloudOS – importing tools and pipelines on the Cloud

Linking your account



Link private repositories



Access and use scripts and pipelines instantly



Share repos with collaborators in your workspace

Import and run a pipeline

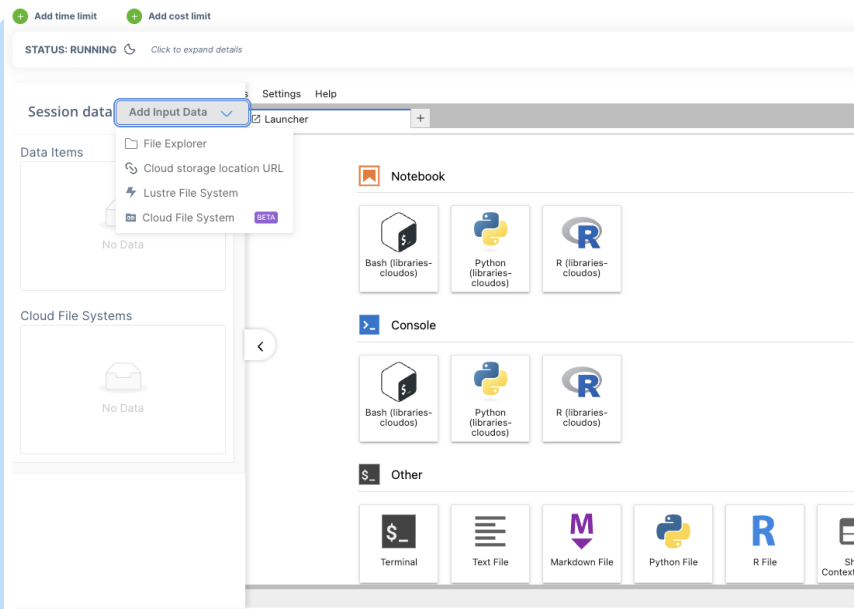


Easy to Setup and Configure (no command line needed)

Monitor, Clone & Resume analyses

Share & collaborate together

Run a pipeline interactively



Use `git clone` to pull your repo

Test and edit your Pipelines

Install and manage dependencies using
Conda, Podman and Quay.io

Scale compute resources to your analysis

Add Data to your Workspace

Upload from the VDI

Import from S3

Link an S3 Bucket

HPC files

Web Browser files

1000 Genomes

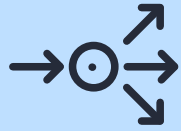
GNOMAD

Lifebit Datasets

AWS Public Resources

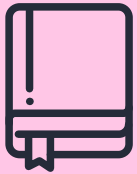
Your own bucket

CloudOS Resources



Nextflow Template

<https://github.com/lifebit-ai/template4users>



Documentation

<https://lifebit.atlassian.net/wiki/spaces/CD/overview?homepageId=168231575>



Help Desk

CloudOS demo

Dashboard ⓘ

+ New

New Cohort

Cohort (62) [View](#) 

Cohort name	Owner	Date created	Date modified	Number of participants	Datasource
Sam_ancestry	Sam Tallman	01/12/2026 16:54:30	01/13/2026 12:51:20	90173	source_data_main_programme_v19
EAS_high	Sam Tallman	01/12/2026 16:43:54	01/13/2026 12:21:02	90173	source_data_main_programme_v19
ancpack_test	Sam Tallman	01/12/2026 16:31:25	01/12/2026 16:41:56	396	source_data_main_programme_v19
vlad test	Vlad Dembrovskiy	11/13/2025 18:16:34	11/13/2025 18:24:36	15189	source_data_main_programme_v19
workflows_demo	Emily Perry	11/11/2025 13:44:01	11/12/2025 09:05:57	90177	omop_data_main_programme_v19
source_demo_w...	Emily Perry	10/23/2025 09:52:01	10/23/2025 10:04:50	459	source_data_main_programme_v19
omop_demo_wor...	Emily Perry	10/23/2025 09:49:41	10/23/2025 09:51:23	67	omop_data_main_programme_v19
cloduos_demo	Emily Perry	10/17/2025 09:03:13	10/17/2025 09:07:04	38860	source_data_main_programme_v19
mvf_svwf_test_c...	Matthieu Vizuite-Forster	09/12/2025 10:30:49	09/12/2025 10:30:49	88491	source_data_100kv17_covidv5
augusto explores	Augusto Rendon	07/31/2025 11:53:23	11/17/2025 17:28:21	90173	source_data_main_programme_v19

Interactive analyses (272) [View all](#)

New Analysis

Status	Session name	Owner	Project	Created at	Total Running time	Last time saved	Cost	Resources
--------	--------------	-------	---------	------------	--------------------	-----------------	------	-----------

Trash

Applications Places System    Lifebit FedPaaS — Mozill...

9. Software licensing requirements

Responsibility



- Check for academic-only licenses (software and data)
- Purchase your own personal licenses
- We are not responsible if you use software or data you don't have a license for

10. Getting help and questions

Getting help



Check our documentation:
<https://re-docs.genomicsengland.co.uk/>
Click on the documentation icon in the environment



Contact our Service Desk:
<https://jiraservicedesk.extge.co.uk/plugins/servlet/desk>

Training sessions

3rd Tuesday every month

Introduction to the RE

17/3

21/4

19/5



Materials
from past
training all
online

Training sessions

10/3

Working with the new aggregate VCFs – AggV3

14/4

Building cancer cohorts

12/5

Building rare disease cohorts



Materials from
past training
all online

Feedback



Thank you

Visit: <https://re-docs.genomicsengland.co.uk/>